

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:14:33 ; Search time 56.06 Seconds
(without alignments)
462.792 Million cell updates/sec

Title:

US-09-823-676-2

Perfect score:

1392

Sequence:

1 MELNRSEADEAKAETPTGG.....IYGMSSGGGGGGGATRPAPF 270

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392	100.0	270	2 T06118	hypothetical prote
2	706.5	50.8	311	2 F86341	hypothetical prote
3	646.5	46.4	292	2 C71448	hypothetical prote
4	645.5	46.4	302	2 H96792	unknown protein F1
5	629.5	45.2	317	2 D84890	probable AT-hook D
6	625.5	44.9	339	2 T06612	hypothetical prote
7	616	44.3	285	2 E84766	probable AT-hook D
8	605	43.5	324	2 T04576	hypothetical prote
9	581	41.7	265	2 T47898	hypothetical prote
10	572	41.1	310	2 T47695	hypothetical prote
11	554	39.8	257	2 B84860	hypothetical prote
12	417	30.0	206	2 G86279	hypothetical prote
13	274	19.7	296	2 G84772	F14L17.27 protein
14	255.5	18.4	334	2 T04572	AT-hook DNA-bindin
15	246.5	17.7	347	2 T06584	hypothetical prote
16	243	17.5	439	2 T52291	probable DNA-bind
17	239.5	17.2	404	2 T05553	hypothetical prote
18	239	17.2	455	2 T05081	hypothetical prote
19	227	16.3	365	2 T06615	hypothetical prote
20	225	16.2	348	2 T02462	hypothetical prote
21	218	15.7	345	2 T01348	probable AT-hook D
22	211.5	15.2	348	2 T47923	hypothetical prote
23	162.5	11.7	574	2 A84782	probable DNA-bind
24	146	10.5	405	2 T29167	hypothetical prote
25	139	10.0	307	2 T27609	hypothetical prote
26	137	9.8	694	2 F70868	hypothetical prote
27	135.5	9.7	900	2 T23416	hypothetical glyci
28	134.5	9.7	839	2 F75518	hypothetical prote
29	131	9.4	388	2 T29173	hypothetical prote

30	131	9.4	1381	2 E70806	hypothetical glyci
31	129.5	9.3	575	2 S5327	protein kinase egg
32	129.5	9.3	1844	2 T51890	related to Nup98-N
33	129.5	9.3	13288	2 T03099	mucin, submaxillar
34	128.5	9.2	1585	2 T31611	hypothetical prote
35	128	9.2	2468	2 A83412	hypothetical prote
36	125.5	9.0	2038	2 A43742	female sterile hom
37	125	9.0	528	2 D46449	hypothetical prote
38	124.5	8.9	867	2 S57795	probable deoxyribo
39	124	8.9	1655	2 T13998	gene mastermind pr
40	124	8.9	3190	2 T13828	CREB-binding prote
41	122	8.8	615	2 H70589	hypothetical glyci
42	122	8.8	979	2 A35913	regulatory factor
43	121.5	8.7	670	2 T49510	fibroin-3 related
44	121.5	8.7	1231	2 S30185	insulin receptor s
45	121	8.7	418	2 B64924	hypothetical prote

ALIGNMENTS

RESULT 1

T06118

hypothetical protein F23E12.50 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000

C:Accession: T06118

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Hohels

submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15485

A:Accession: T06118

A:Molecule type: DNA

A:Residues: 1-270 <BEV>

A:Cross-references: EMBL:AL022604; GSPDB:GN00062; ATSP:F23E12.50

A:Experimental source: cultivar Columbia; BAC clone F23E12

C:Genetics:

A:Gene: ATSP:F23E12.50

A:Map position: 4

C:Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match	100.0%	Score 1392;	DB 2;	Length 270;
Best Local Similarity	100.0%;	Pred. No. 6.6e-85;		
Matches 270;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MELNRSEADEAKAETPTGGATSSATASGSSRRRGRPGAGSKNPKPPTIITRDSNV 60		
Db	1	MELNRSEADEAKAETPTGGATSSATASGSSRRRGRPGAGSKNPKPPTIITRDSNV 60		
QY	61	LRSHVLEVTSGDISEAVSTYATRRGCGVCIIISCTGAVTNVTIRQAPAGGGVITLHGR 120		
Db	61	LRSHVLEVTSGDISEAVSTYATRRGCGVCIIISCTGAVTNVTIRQAPAGGGVITLHGR 120		
QY	121	FDILSLTGALPPAPPGAGGLTVYLAGGGQVVGNGVAGSLIASGPVVLMAASFANAY 180		
Db	121	FDILSLTGALPPAPPGAGGLTVYLAGGGQVVGNGVAGSLIASGPVVLMAASFANAY 180		
QY	181	DRLPIEEETPPRTTGVOQQQPEASQSSSEVTGSGAQACESNLQGGGGGVAFYNLGMN 240		
Db	181	DRLPIEEETPPRTTGVOQQQPEASQSSSEVTGSGAQACESNLQGGGGGVAFYNLGMN 240		
QY	241	MNNFQSGGDIYGMSSGGGGGGGATRPAPF 270		
Db	241	MNNFQSGGDIYGMSSGGGGGGGATRPAPF 270		

RESULT 2

F86341

hypothetical protein F9H16.12 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C:Accession: F86341

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

[illegible]

Db 276 QL---PVEGWPGNSGGRG 290

H96792
unknown protein F14G6.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: H96792
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; D
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.;
Rizzo, M.; Rooney, S.; Rowley D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H96792
A:Status: preliminary
A:Molecule type: DNA

```
C;Genetics:
A;Gene: F14G6.10
A;Map position: 1
C;Superfamily: Arabidopsis thaliana hypothetical protein Tl2H17.200

Query Match      46.4%   Score 645.5; DB 2; Length 302;
Best Local Similarity 49.5%; Pred. No. 1.2e-35;
Matches 138; Conservative 33; Mismatches 63; Indels 45; Gaps

Qy    9 DEAKETPTGGATSSATASGSSRRRPRGPAGSKNKPPTIITROSPNVLRSHVLEV 68
       || : | : | : |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    52 DESDNKDP-----GSDPVTSGTGKRGRPPRGPSKNPKPPIVTRDSPNVLRSHVLEV 106

Qy    69 TSGSDISEAVTYATRRCGVCIIISGTCAVNVTIRQPAPAG-----GGVITHGRF 121
       |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    107 SSGADIVESVTYARRRGVGSILSNGCTVANSLRQPATTAAHGANGTGGVVALHGRF 166

Qy    122 DILSLTGALPPPAPPAGGITTVYLAGQGQQVGNGVAGSLIASGPVLMAASFANAYD 181
       |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    167 EILSLTGVLPPPAPPGGGLSIFLSGVQGVIGNVVAPLVSAGFVILMAASFNNATFE 226

Qy    182 RLPTEEEETPPRTTGVQQQPEASQSSEVTGSGAQACESNLQGNNG -----GG 231
       ||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

QY 232 VAFNLGMNMNNFQSGGDIYMSGGSGGGGATRPAP 270
| : : | : : |
|||

FT DOMAIN 480 709 9.68; Score 133; DB 1; Length 709;
FT DOMAIN 3 6 LIGAND-BINDING.
FT DOMAIN 182 187 POLY-GLN.
FT DOMAIN 201 207 POLY-PRO.
FT DOMAIN 254 262 POLY-ALA.
SQ SEQUENCE 709 AA; 77391 MW; 40E7666137B97B6B CRC64;
Query Match
Best Local Similarity 24.68; Pred. No. 0.15;
Matches 70; Conservative 32; Mismatches 126; Indels 56; Gaps 11;
Qy 25 ATAGSSSG--RRRGRPGAGKPKPTTITRDSNVLRSV-LEVTSGSDISEAVSTY 81
Db 8 AATGSSCGARRRSGASTSKDYLSTSVISDAKELCKANVSGLGLVEALEHLSSG 67
Qy 82 ATRRG-CGVCISGTGAVNTVIRQAPAGGGVITLHGRFDILSLGTALPPPPAG 140
Db 68 EQLRGDCMYAPLLGGPPVVRTPCLPLVECKGSL-----DDGPGKGTEETAEYTPFKG 121
Qy 141 GLTVYLAGGQGVVGNVAGSLIAGPVL-----MAAFANAVYDRLPDIE- 186
Db 122 G---YNRKLEAESLGCSGAGSGTLELSTLSLYKSGTLDAAAYQTDYVNFPLAL 178
Qy 187 EETPPRTTGVQOQPEASOSSEVTGSGAQACESNLQGGNGGGV-----A 233
Db 179 AGQPPPHPRKLENP-LDYGSAWAAAAOCRYGDLASLHGGGAGPGSGPSTAASSS 237
Qy 234 FYNGLMNNNFQSGDGIYMGSGGGGGGA-----TRP 268
Db 238 WHTL-----FTTERGQLYGLCGGGGGGPGGAGAVPYGYTRP 275
RESULT 2
APMU_PIG STANDARD; PRT; 1150 AA.
ID APMU_PIG
AC P12021;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Apomucin (Mucin core protein) (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Submaxillary gland;
RX MEDLINE=91236743; PubMed=2033060;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Zhao Y., Hill R.L.;
RT "Porcine submaxillary mucin contains a cysteine-rich,
RT carboxyl-terminal domain in addition to a highly repetitive,
RT glycosylated domain.";
RL J. Biol. Chem. 266:9678-9686(1991).
RN [2]
RP SEQUENCE OF 1-503 FROM N.A.
RC TISSUE-Submaxillary gland;
RX MEDLINE=88087170; PubMed=2826455;
RA Timpte C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
RT "Porcine submaxillary gland apomucin contains tandemly repeated,
RT identical sequences of 81 residues.";
RL J. Biol. Chem. 263:1081-1088(1988).
RN [3]
RP SEQUENCE OF 45-80.
RC TISSUE-Submaxillary gland;
RX MEDLINE=87280230; PubMed=3611111;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Tounadje A.,
RA Johnson W.C. Jr., Hill R.L.;
RT "Structural properties of porcine submaxillary gland apomucin.";
RL J. Biol. Chem. 262:11339-11344(1987).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.
RC TISSUE-Submaxillary gland;
RX MEDLINE=97248516; PubMed=9092502;
RA Gerken T.A., Owens C.L., Pasumarthy M.;
RT "Determination of the site-specific O-glycosylation pattern of the
RT porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
RT for the polypeptide:galnac transferase binding site.";
RL J. Biol. Chem. 272:9709-9719(1997).
CC -1- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
CC SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY
CC GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
CC THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
CC ENVIRONMENT.
CC -1- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
CC MULTIMERIC MUCIN STRUCTURE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -1- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
CC RESIDUES.
CC -1- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
CC THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
CC RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE
CC GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE. THE
CC PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
CC ENHANCE GLYCOSYLATION.
CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M61883; AAA30998.1; -;
DR EMBL: M21174; AAA30990.1; -;
DR PIR: A40009; A40009.
DR InterPro: IPR000359; Cys_knot.
DR InterPro: IPR001007; WFEC.
DR Pfam: PF00007; Cys_knot; 1.
DR SMART: SM00041; Ct; 1.
DR SMART: SM00214; WVC; 1.
DR PROSITE: PS01185; CTCK_1; 1.
DR PROSITE: PS01225; CTCK_2; 1.
DR PROSITE: PS01208; WFEC; 1.
KW Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN <1 368 81 AA TANDEM REPEATS.
FT REPEAT <1 44 1.
FT REPEAT 45 125 2.
FT REPEAT 126 206 3.
FT REPEAT 207 287 4.
FT REPEAT 288 368 5.
FT REPEAT 369 391 6 (INCOMPLETE).
FT REPEAT 929 995 WFEC.
FT DOMAIN 1062 1146 CTCK.
FT DISULFID 1062 1109 BY SIMILARITY.
FT DISULFID 1076 1123 BY SIMILARITY.
FT DISULFID 1085 1139 BY SIMILARITY.
FT DISULFID 1089 1141 BY SIMILARITY.
FT DISULFID ? 1145 BY SIMILARITY.
FT CARBOHYD 46 46 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 50 50 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 51 51 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 57 57 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 58 58 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 61 61 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 66 66 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 67 67 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 73 73 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 74 74 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 76 76 O-LINKED (GALNAC. .) (MUCIN TYPE).
FT CARBOHYD 77 77 O-LINKED (GALNAC. .) (MUCIN TYPE).

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:21:23 ; Search time 49.21 Seconds
(without alignments)
212.442 Million cell updates/sec

Title: US-09-823-676-2

Perfect score: 1392

Sequence: 1 MELNRSDEAKAETPTGG.....IYMGSGGGGGGATPAF 270

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	133	9.6	709	1 ANDR_RABIT	P49699 oryctolagus
2	128.5	9.2	1150	1 APMU_PIG	P12021 sus scrofa
3	125.5	9.0	2038	1 FSH_DROME	P13709 drosophila
4	123.5	8.9	515	1 MEF2_DROME	P40791 drosophila
5	123.5	8.9	884	1 ANDR_EULFC	O97776 eulemur ful
6	122.5	8.8	895	1 ANDR_MACEA	O97952 macaca fasc
7	122	8.8	979	1 RFX1_HUMAN	P22670 homo sapien
8	121.5	8.7	895	1 ANDR_PAPHA	O97960 papio hamad
9	121.5	8.7	911	1 ANDR_PANTR	O97775 pan troglod
10	118.5	8.5	919	1 ANDR_HUMAN	P10275 homo sapien
11	117	8.4	367	1 BET3_MESAU	O90029 mesocricetu
12	117	8.4	1174	1 CIKE_DROME	O02280 drosophila
13	116.5	8.4	1233	1 IRS1_MOUSE	P35569 mus musculus
14	116	8.3	1322	1 SUS_DROME	P22293 drosophila
15	115.5	8.3	378	1 CSP_PLACB	P08672 plasmodium
16	115	8.3	603	1 YP25_MYCTU	Q10637 mycobacteri
17	115	8.3	1672	1 PMPB_CHLMU	O96372 chlamydia m
18	113.5	8.2	553	1 FNX1_HUMAN	Q12348 homo sapien
19	113.5	8.2	1319	1 MNL_HUMAN	Q10571 homo sapien
20	113.5	8.2	1901	1 YZ08_MYCTU	O53553 mycobacteri
21	113.5	8.2	2090	1 N214_HUMAN	P35658 homo sapien
22	113	8.1	801	1 Y747_MYCTU	O53810 mycobacteri
23	113	8.1	2333	1 PGCA_CANPA	Q28343 canis famil
24	112.5	8.1	1235	1 IRS1_RAT	P35570 rattus norv
25	112	8.0	543	1 YP91_MYCTU	Q50630 mycobacteri
26	110.5	7.9	654	1 TFE2_HUMAN	P15923 homo sapien
27	110.5	7.9	860	1 EUS_MOUSE	P54320 mus musculus
28	110.5	7.9	1156	1 GLH4_CAEEL	O96743 caenorhabdi
29	110.5	7.9	2517	1 NCR2_HUMAN	O97618 h nuclear r
30	109.5	7.9	589	1 DYBB_MOUSE	O92188 mus musculus
31	109	7.8	672	1 PHX5_MOUSE	P08399 mus musculus
32	109	7.8	732	1 TAU_MOUSE	P10637 mus musculus
33	109	7.8	825	1 ICP0_HSV2H	P26284 herpes simp

ALIGNMENTS

RESULT_1	ANDR_RABIT	STANDARD	PRT	709 AA
AC	P49699;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Androgen receptor (Dihydrotestosterone receptor) (Fragment).			
GN	AR OR NR3C4.			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_TaxID=9986;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NEW ZEALAND WHITE; TISSUE-Prostate;			
RX	MEDLINE=96044663; PubMed=7559153;			
RA	Kronrad A., Wilson J.D., McPhaul M.J.;			
RT	"Cloning and partial sequence of the rabbit androgen receptor:			
RT	expression in fetal urogenital tissues.";			
RL	J. Androl. 16:209-212(1995).			
CC	-!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.			
CC	-!- SUBCELLULAR LOCATION: Nuclear.			
CC	-!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN, A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.			
CC	-!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY. NR3 SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: U16366; AAC48469.1; .			
DR	HSSP: P06536; IGDC.			
DR	InterPro: IPR001103; Androgen_recep.			
DR	InterPro: IPR000536; Hormone_rec_lig.			
DR	InterPro: IPR001628; zf-C4.			
DR	Pfam: PF02166; Androgen_recep; 1.			
DR	Pfam: PF00104; hormone_rec; 1.			
DR	Pfam: PF00105; zf-C4; 1.			
DR	SMART: SM00430; HOL1; 1.			
DR	SMART: SM00399; ZNF_C4; 1.			
DR	PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.			
KW	Receptor; Transcription regulation; DNA-binding; Nuclear protein;			
KW	Zinc-finger; Steroid-binding.			
FT	NON_TER 1			
FT	DOMAIN <1 347			
FT	DNA_BIND 349 414			
FT	ZN_FING 349 369			
FT	ZN_FING 385 409			

Q9j1l3 rattus norv
P09026 mus musculus
Q99372 rattus norv
O06794 mycobacteri
P27921 gallus gall
P21519 drosophila
P49657 drosophila
P10636 homo sapien
Q61982 mus musculus
P23490 homo sapien
P19622 homo sapien
P03181 epstein-bar

34 109 7.8 910 1 ILF3_RAT
35 108.5 7.8 433 1 HXB3_MOUSE
36 108.5 7.8 864 1 ELS_RAT
37 108.5 7.8 914 1 WA22_MYCTU
38 108 7.8 323 1 JUND_CHICK
39 108 7.8 1596 1 MAM_DROME
40 107.5 7.7 843 1 MNB_DROME
41 107 7.7 757 1 TAU_HUMAN
42 107 7.7 2318 1 NTC3_MOUSE
43 106.5 7.7 316 1 LORI_HUMAN
44 106.5 7.7 332 1 HME2_HUMAN
45 106.5 7.7 660 1 YHL1_EBV

Thu Jul 11 11:01:05 2002

QY 84 RRGCGVCIISGTGAVNTIROPAPAGGGVITLHGRFDILSLGTALPPAP---PGAG 140
Db 173 QGSLAICVLCANGVSSVTLRQP--DSSGGVLTVEGRFEILSLSGTFMPSDSDGTRSR TG 230
QY 141 GLTVYLAGOGQGVGNGVAGSLIAGSPVVLMAASFANAVYDRLPIEEETPPP 193
Db 231 GMSVSLASPDGRVGGVAGLVAAATPIQVVVGTFLGGTN-----QOEQT PKP 278

RESULT 15

T06584
probable DNA-binding protein - garden pea
C:Species: Pisum sativum (garden pea)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 01-Dec-2000
C:Accession: T06584; T06582
R:Sato, N.; Kazuno, A.A.; Ohta, N.; Ohshima, K.
submitted to the EMBL Data Library, June 1996
A:Description: Identification of a novel family of DNA-binding proteins with two AT-hook
A:Reference number: Z15774
A:Accession: T06584
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-347 <SAT>
A:Cross-references: EMBL:X98738; PIDN:CAA67290.1
A:Accession: T06582
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-114,'S',116-334 <SA2>
A:Cross-references: EMBL:X98739; PIDN:CAA67291.1
A:Experimental source: cv. Alaska
C:Keywords: DNA binding

Query Match 17.7%; Score 246.5; DB 2; Length 347;
Best Local Similarity 36.7%; Pred. No. 2.8e-09;
Matches 61; Conservative 30; Mismatches 66; Indels 9; Gaps 5;
QY 16 TPTGATSSATASGSSGRPRGPRGSKNKPPTI---ITRDSNVLRSHVLEVTSGS 72
Db 115 TPMSAPANSTODSTPSEKRGGRGPRGSKQQLAALGDWMTSSAGLAFSPHVTIAAGE 174
QY 73 DISEAVSTYATRRGCGVCIISGTGAVNTIROPAPAGGGVITLHGRFDILSLGTAL- 131
Db 175 DIAAKLLLSQQRPRALCILSGTGIAKSKVTLRQPASTNAG--VTEGKFOILSLSGSYLV 232
QY 132 -PPPAPPG-AGGLVYLAGOGQGVGNGVAGSLIAGSPVVLMAASF 175
Db 233 SEDGGTNRGTGGISVLSRDRGHVIGGSA-MLIAGSPIQLVVCSE 277

Search completed: July 10, 2002, 22:26:43
Job time: 730 sec

Query Match 39.8%; Score 554; DB 2; Length 257;
 Best Local Similarity 52.8%; Pred. No. 1.1e-29;
 Matches 113; Conservative 30; Mismatches 37; Indels 34; Gaps 4;

QY 16 TPTGGATSS-----ATAGSSS-----GRRPRGRPAGSKNPK 48
 DB 8 TTVSGSKVPMRNHEATRGNTNNLRALPKAVQPVSSIEGEMAKRPRGRPAGSKNPK 67
 QY 49 PPTIITRDSPNVLRSHVLEVTSGSDISAVSTYATRRCGVCIIISGTGAVTNVITRQPA 108
 DB 68 PPIIVTHDSFNLRAVAVSSIGDICTSLDFARRKQGLCILSANGCVTNVTLRQPA- 126
 QY 109 PAGGCVITLHGRFDILSTGTALPPAPPGAGGLTVYLAGGQGVGVGNVAGSLIAGSPV 168
 DB 127 -SSGAIVTLHGRYELSLGLSTLPPAPLGLITGLIYLAGGQGVGVGVGLIAGSPV 185
 QY 169 VLMASAFANAVYDRLPIEBEETPPRTTGVQOQQ 202
 DB 186 VLMASAFNVAVDRLPMDDE-----AASMQNOQ 214

RESULT 12
 G86279
 F14L17.27 protein - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001
 C:Accession: G86279
 R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: G86279
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-206 <STO>
 A:Cross-references: GB:AE005172; NID:g7262692; PIDN:AAF43950.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 30.0%; Score 417; DB 2; Length 206;
 Best Local Similarity 46.0%; Pred. No. 9.1e-21;
 Matches 91; Conservative 32; Mismatches 59; Indels 16; Gaps 5;

QY 35 RPRGRPAGSKNPKPTIITRDSPNVLRSHVLEVTSGSDISAVSTYATRRCGVCIIISG 94
 DB 6 RPRGRPAGSKNPKAPFTVIDPP--MSPIILEVSGNDVVEALNFCRGAIGFCVLVG 63
 QY 95 TGAVTNVTIROPAPAGGCVITLHGRFDILSTGTALPP-----PAPPAGGLTVYLAG 149
 DB 64 SGVADVTLRQSPAPAGSTITFHGKFDLLSVATLPLPPTLSLSPVSNFTVSLAGP 123
 QY 150 QGVGVGNVAGSLIAGSPVLMASAFANAVYDRLPIEBEETPPRTTGVQOQQPASOSS 209
 DB 124 QGVIGGFVAGPLVAVGTIVFATSKFNPVHRLPATEEE---QRNSA---EGEEGOSP 177

QY 210 EYTGGAQACENLOGGN 227
 DB 178 PVSAGGG---ESMYVGG 192

RESULT 13
 G84747
 Ar-hook DNA-binding protein (AHP1) [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: G84747
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: G84747
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-296 <STO>
 A:Cross-references: GB:AE002093; NID:g2459442; PIDN:AAB80677.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g33620
 A:Map position: 2

Query Match 19.7%; Score 274; DB 2; Length 296;
 Best Local Similarity 35.1%; Pred. No. 3.6e-11;
 Matches 87; Conservative 28; Mismatches 81; Indels 52; Gaps 11;

QY 22 TSSATAGSSSSGRRPRGRPAGSKNPKPTIITRDSPNVLR-----HVLEVTGSDI 74
 DB 70 TVSQPSGGDGEKKRPPGSSK-----RLQLALGSTGIGFTPHVLTVLAGEDV 121
 QY 75 SEAVSTYATRRCGVCIIISGTGAVTNVITRQPAAPAGGCVITLHGRFDILSTGT--ALP 132
 DB 122 SSKIMALTHNGPRAVCVLSANGAISNVLQSA--TSGTIVYEGFELSLSGSFHLE 179
 QY 133 PPAPPG-AGLVYLAGGQGVGVGNVAGSLIASPVLMAAFANAVYDRLPIEBEE-- 189
 DB 180 NNGQRRTGGLSVLSLSSPDGVLGGSVAGLLIAASPVQIVVGSF-----LPDGEKEPK 232
 QY 190 -----TTP--PRTTGVQOQQPASOSSSEVTGSGAQACENLOGGNGG-----GGVA 233
 DB 233 QHVQGMGLSSPVLPRVAPTOVLTMTSPSQSRGTMS-----ESSCGGGHGHQHTGGP- 286
 QY 234 FYNLGMN 241
 DB 287 -YNNNTNM 293

RESULT 14
 T04572
 hypothetical protein T12H17.160 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
 C:Accession: T04572
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Hoyer, A.; Brandt, A.; Bancro
 submitted to the Protein Sequence Database, February 1998
 A:Reference number: Z15377
 A:Accession: T04572
 A:Molecule type: DNA
 A:Residues: 1-334 <BEV>
 A:Cross-references: EMBL:AL021635
 A:Experimental source: Cultivar Columbia; BAC clone T12H17
 C:Genetics:
 A:Map position: 4
 A:Introns: 139/1; 161/3; 205/3
 A:Note: T12H17.160

Query Match 18.4%; Score 255.5; DB 2; Length 334;
 Best Local Similarity 27.5%; Pred. No. 6.8e-10;
 Matches 64; Conservative 38; Mismatches 68; Indels 63; Gaps 6;

QY 17 PTGGATSSATAGSSSG--RRPRGRP-----40
 DB 53 PQNSFTPSAAMDGFSGPIKKRRGRPRKYGDGAAYTLSPNPISAAPTTSHVIDFSTTS 112
 QY 41 -AGSKNPKPPTIITRDSPNV-----LRSHVLEVTSGSDISPAVSTYAT 83
 DB 113 EKRGMKAPPTPTSSFRPKYQVENLGENWSPSAAANFTPHIITVNAVAGEDVTKRIISFQ 172

R:Bevan, M.; Hilbert, H.; Braun, M.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, February 1998
A:Reference number: Z15377
A:Accession: T04576
A:Molecule type: DNA
A:Residues: 1-324 <BEV>
A:Cross-references: EMBL:AL021635
A:Experimental source: cultivar Columbia; BAC clone T12H17
C:Genetics:
A:Map position: 4
A:Note: T12H17.200
C:Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 43.5%; Score 605; DB 2; Length 324;
Best Local Similarity 56.8%; Pred. No. 6e-33;
Matches 129; Conservative 26; Mismatches 42; Indels 30; Gaps 6;

QY 19 GGATSSATASGSSG-----RRPRGRPAGSKNPKPPTITRDSNVLRSHVLEVTSGSD 73
Db 88 GG-----SEGEGGGGGDHQWTRRRPRGRPAGSKNPKPPTITRDSANALRTHWIEIGDCD 144
QY 74 ISEAVSYATRRCGVCIIISGTGAVNTVIRQAAA-PAGGGVITLHGRFDILSLTCTALP 132
Db 145 LVESVATFARRRQRGVCVMSGTGNTVTIRQPGSHPSGVSVSLHGRFEILSLSGFLP 204
QY 133 PPAPPGAGGLTVYLAGGGGVGVGNVAGSLIAGSPVVLMAASFANAVYDRLPIEEEE--T 190
Db 205 PPAPPTATGUSVYLAGGGGVGVGNVAGSLIAGSPVVLMAASFANAVYDRLPIEEDEMOT 264
QY 191 P-----PPRTTGVQQQOPEASQSSEVTGSGAQACESNLQG 225
Db 265 PVHGGGGGSLSPMPMQQLHQQQAM-----SGHQGLPPLL 304

RESULT 9
T47898
hypothetical protein T4C21.280 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
C:Accession: T47898
R:Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
w.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24479
A:Accession: T47898
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-265 <CHO>
A:Cross-references: EMBL:AL162295
A:Experimental source: cultivar Columbia; BAC clone T4C21
C:Genetics:
A:Map position: 3
A:Note: T4C21.280
C:Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 41.7%; Score 581; DB 2; Length 265;
Best Local Similarity 50.2%; Pred. No. 1.9e-31;
Matches 119; Conservative 31; Mismatches 57; Indels 30; Gaps 3;

QY 34 RRPRGRPAGSKNPKPPTITRDSNVLRSHVLEVTSGDISEAVSYATRRCGVCIIIS 93
Db 59 RRPRGRPAGSKNPKPPTITRDSANAFRCHVMEITNACDVMSLAVFARRRQRGVCVLT 118
QY 94 GTGAVNTVIRQAPAPAGGGVITLHGRFDILSLTCTALPPPPAGGGLTVYLAGGGGV 153
Db 119 GNCVNTVIRQAP-----GGVVVSLHGRFEILSLSGFLPAPPAAASGLKYLAGGQQV 174
QY 154 VGNVAGSLIAGSPVVLMAASFANAVYDRLPIEEEEPPPTTGVQQQOPEASQSSEVTG 213
Db 175 IGSVVGLTASSPVVVMMAASFANAVYDRLPIEEEE-----ETEREIDG 218

QY 214 SGAQACESNLQGGGGGVAFYNLGMNMFQSGGDIYMGSGSGGGGGGATRRAP 270
Db 219 NAARAIGTQKQLMQDATSFIGSPSLNLSVLPGEAY-----WGTORPSF 265

RESULT 10
T47695
hypothetical protein T22E16.220 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Jun-2000
C:Accession: T47695
R:Benes, V.; Wurmback, E.; Drzonek, H.; Ansoerge, W.; Mewes, H.W.; Lemcke, K.; Mayer, submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24472
A:Accession: T47695
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-310 <BEN>
A:Cross-references: EMBL:AL132975
A:Experimental source: cultivar Columbia; BAC clone T22E16
C:Genetics:
A:Map position: 3
A:Note: T22E16.220
C:Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 41.1%; Score 572; DB 2; Length 310;
Best Local Similarity 47.8%; Pred. No. 8.6e-31;
Matches 120; Conservative 44; Mismatches 69; Indels 18; Gaps 6;

QY 5 READEAKAETPTGTGATSSATASGSSGRRPRGRPAGSKNPKPPTITRDSNVLRSH 64
Db 63 QSOEEQNSRDEQP-----AVEPGSGSG-STGRRPRGRPAGSKNPKSPVVVTRKSPNSLQSH 118
QY 65 VLEVTSGSDISEAVSYATRRCGVCIIISGTGAVNTVIRQAPAGGGVITLHGRFDIL 124
Db 119 VLEIATGADVAESLNAFARRRGRGVSLGSLVNTVLRQPA--ASGGVSLRQGFIL 176
QY 125 SITGTALPPP-APPAGGLTVYLAGGGGVGVGNVAGSLIAGSPVVLMAASFANAVYDRL 183
Db 177 SMCGLPITGSGPAAAGLTIYLAGAGGGGVGVGNVAGSLIAGSPVIVIANFCNATYERL 236
QY 184 PTIEETPTPTTGTGVQQQOPEASQSSEVTGSGAQACESNLQGGGGGVAFYNLGMNM-- 241
Db 237 PTIEEQ-----QOEQLQLEDGKKQKEENDDNESGNGSQMPMYNMPNFTP 287
QY 242 NNFQSGGDIY 252
Db 288 NGHQAQHDVY 298

RESULT 11
B84860
hypothetical protein At2g42940 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84860
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: B84860
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <STO>
A:Cross-references: GB:AE002093; NID:g4512661; PIDN:AAD21715.1; GSPDB:GN001139
C:Genetics:
A:Gene: At2g42940
A:Map position: 2

QY 19 GGATSSATASGSSG-----RRPRGRPAGSKNPKPPTITRDSNVLRSHVLEVTSGSD 73
Db 88 GG-----SEGEGGGGGDHQWTRRRPRGRPAGSKNPKPPTITRDSANALRTHWIEIGDCD 144
QY 74 ISEAVSYATRRCGVCIIISGTGAVNTVIRQAAA-PAGGGVITLHGRFDILSLTCTALP 132
Db 145 LVESVATFARRRQRGVCVMSGTGNTVTIRQPGSHPSGVSVSLHGRFEILSLSGFLP 204
QY 133 PPAPPGAGGLTVYLAGGGGVGVGNVAGSLIAGSPVVLMAASFANAVYDRLPIEEEE--T 190
Db 205 PPAPPTATGUSVYLAGGGGVGVGNVAGSLIAGSPVVLMAASFANAVYDRLPIEEDEMOT 264
QY 191 P-----PPRTTGVQQQOPEASQSSEVTGSGAQACESNLQG 225
Db 265 PVHGGGGGSLSPMPMQQLHQQQAM-----SGHQGLPPLL 304

RESULT 9
T47898
hypothetical protein T4C21.280 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
C:Accession: T47898
R:Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
w.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24479
A:Accession: T47898
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-265 <CHO>
A:Cross-references: EMBL:AL162295
A:Experimental source: cultivar Columbia; BAC clone T4C21
C:Genetics:
A:Map position: 3
A:Note: T4C21.280
C:Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 41.7%; Score 581; DB 2; Length 265;
Best Local Similarity 50.2%; Pred. No. 1.9e-31;
Matches 119; Conservative 31; Mismatches 57; Indels 30; Gaps 3;

QY 34 RRPRGRPAGSKNPKPPTITRDSNVLRSHVLEVTSGDISEAVSYATRRCGVCIIIS 93
Db 59 RRPRGRPAGSKNPKPPTITRDSANAFRCHVMEITNACDVMSLAVFARRRQRGVCVLT 118
QY 94 GTGAVNTVIRQAPAPAGGGVITLHGRFDILSLTCTALPPPPAGGGLTVYLAGGGGV 153
Db 119 GNCVNTVIRQAP-----GGVVVSLHGRFEILSLSGFLPAPPAAASGLKYLAGGQQV 174
QY 154 VGNVAGSLIAGSPVVLMAASFANAVYDRLPIEEEEPPPTTGVQQQOPEASQSSEVTG 213
Db 175 IGSVVGLTASSPVVVMMAASFANAVYDRLPIEEEE-----ETEREIDG 218

RESULT 14

US-09-219-849-50

; Sequence 50, Application US/09219849

; Patent No. 6150081

; GENERAL INFORMATION:

; APPLICANT: VAN HEERDE, GEORGE V.

; APPLICANT: VAN RIJN, ALEXIS C.

; APPLICANT: BOUNSTRA, JAN B.

; APPLICANT: DE WOLF, FREDERIK A.

; APPLICANT: MOOBROEK, ANDREAS

; APPLICANT: WERTEN, MARC W.T.

; APPLICANT: WIND, RICHELE D.

; APPLICANT: VAN DEN BOSCH, TANJA J.

; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN

; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE

; PREPARATION THEREOF

; FILE REFERENCE: 2728-2

; CURRENT APPLICATION NUMBER: US/09/219,849

; CURRENT FILING DATE: 1998-12-23

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 50

; LENGTH: 595

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic

; OTHER INFORMATION: amino acid sequence

US-09-219-849-50

Query Match

Best Local Similarity 7.3%; Score 101.5; DB 4; Length 595;

Matches 71; Conservative 19; Mismatches 118; Indels 95; Gaps 15;

QY 6 SEADAKAETPTG--GATSSATASGSSRRPRGRPAGSKNKPPTIITRDSNVLR 63

Db 92 SGLDGAAGDAGPAGPKGPGSPGNGAPGQMGPRGLP--GERGRPGP----- 137

QY 64 HVLEVTSGSDISEAVSTYATRRGCGVCIIISGTGAVTNVTIROPAPAG--GGVITLHGRFD 122

Db 138 -----GTAGARGNDGAVGAAGPPGPTG-----PTGPGFPGAVGAKGEAG 177

QY 123 ILSLTGTALP-----PPAPPAGGLTVYLAGGQGVVGNVAGSLIASG--PVVLMASFA 176

Db 178 PGARGSEGPQGVGRGEPGPPAG-----AAGPAGNPGADGQPCAKGANGAPGAGPFP 233

QY 177 NAVYDLPIEEETPP-----PRTTGVOQQOPEASQSSEVTG 213

Db 234 GARGPSGP--QGPSGPPGKNGSGEPGAPGNKGTGAKGEPGATGVQGPAGEEGK--RG 291

QY 214 SGAQACESNL-----QGGNGGGGVAFYNLGMNMNMFQSGDIYMGSGSG--GGGGGAT 266

Db 292 ARGEPPSGLPGPPGERGGPSRG-----FPGAD--GVAGPKGPSGERGAP 335

QY 267 RPA 269

Db 336 GPA 338

RESULT 15

US-09-219-849-49

; Sequence 49, Application US/09219849

; Patent No. 6150081

; GENERAL INFORMATION:

; APPLICANT: VAN HEERDE, GEORGE V.

; APPLICANT: VAN RIJN, ALEXIS C.

; APPLICANT: BOUNSTRA, JAN B.

; APPLICANT: DE WOLF, FREDERIK A.

; APPLICANT: MOOBROEK, ANDREAS

; APPLICANT: WERTEN, MARC W.T.

Db 1425 AAHGHAHRVROGPERVLGGHGVDPYRORRCHAAAGDEGAVAV-GRVDPALAEALVEALVCG 1483
QY 130 ALPPPPAPP-AGGUTVYLA-----GQGVGVGVGNVAGSLIASG 166
Db 1484 LHPVLPQPHGLAGGAVDAVEAHEGLVLLPRVPHLRDDEGHGPGRGAVAGRLADVVLVPE 1543
QY 167 PVV-LMAASFANAVYDR-----LPIEEETPPPTTGVQOOOPEASOSSEVTVGSGAQA 218
Db 1544 PLAGVPGAAVADAARRVAGAGLPERGEORVPVGRGPGVGHREVVVGGAAALPARPGG 1603
QY 219 CESNLOGGNGGG------VAFYNLGMNMNMF 244
Db 1604 LRGRGRGGRGGGGGGGGRGPRGGRGRRRRRWRPCAGEWGAGPDSVFFFSLG----- 1657
QY 245 QFSGGDIYMGSGSGGGGGGATR 267
Db 1658 ---GGRGRGGRGGRGGRAPR 1677

RESULT 12

US-08-195-152-2
; Sequence 2, Application US/08195152
; Patent No. 5679541
; GENERAL INFORMATION:
; APPLICANT: Bonini, Nancy M.
; APPLICANT: Leiserson, William M.
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,152
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59551/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-152-2

Query Match 7.4%; Score 103; DB 1; Length 760;
Best Local Similarity 23.6%; Pred. No. 0.76; Indels 72; Gaps 11;
Matches 64; Conservative 28; Mismatches 107
QY 19 GGATSSATASGSSGRRPRGPKGKPKPTTIITRDS-----PNVLRSHVLE 67
Db 91 GGGGGGGVGGVGCSPGLPLP--PQSQPLQPTIGSLASLSCHYNGNANPNVNSSCSL 148
QY 68 VTSGSDISEAVSTYATRCGCVCI--ISGTGAVNTVITRQPAAPAGGCVTLHGRFDILS 125

Db 149 ATASSFQAQSAGSSESYQQAGGTSGGSGGEDGV-----VGGATVMSHWHTHDGTG 197
QY 126 LTGTALPPPPAPPAGGGLTVYLAGGQGVGVGNVAGSLIASGPV-----VLMMAASFAN 177
Db 198 SSAARKSESSESPG----QVHASLDSNGSVAGSNLYGCSASNPLDGGAVAVNSSAVAAAA 253
QY 178 AVYD-----RLPIEEETPPP-----RTTGVOOOOPE-----ASOSSEVTVGSGAQ 217
Db 254 AVYDGGKHDYIYYNMOQYTPPFYSGVGTPTAAATAARQAQAKMEPGAAAAAAAAILTPSYAA 313
QY 218 ACESNLQ-----CGNGGGGVAFYN 236
Db 314 SGNNNSQLSSPYAGYNNFQQDYGG--YYN 342

RESULT 13

US-09-219-849-48
; Sequence 48, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOEBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-219-849-48

Query Match 7.3%; Score 101.5; DB 4; Length 595;
Best Local Similarity 23.4%; Pred. No. 0.74;
Matches 71; Conservative 19; Mismatches 118; Indels 95; Gaps 15;
QY 6 SEADEAKAETTTTG--GATSSATASGSSGRRPRGPKGKPKPTTIITRDSPNVLR 63
Db 92 SGLDGAKGDAGPAGPKGPGSPGENGAGQMGPRGLP--GERGRPGP----- 137
QY 64 HVLEVTSGSDISEAVSTYATRCGCVCIISGTGAVNTVITRQPAAPAG--GGVITLHGRFD 122
Db 138 -----GTAGARGNDGAVGAGPPGPTG-----PTGPPGPGAVGAKGEAG 177
QY 123 ILSLTGTALP-----PPAPPAGGLTVYLAGGQGVGVGNVAGSLIASG--PVVLMMAASFA 176
Db 178 PQARGSEGPGQVGRGEPGPPGAG-----AAGPAGNPGADGQPGAKGANGAPGIAGAPGP 233
QY 177 NAVYDLRLPIEEETPP-----PRTTGVOOOOPEASOSSEVTVG 213
Db 234 GARGSPGP--QGPSGPPGPKGNSGEPGAPGNKGTGAKGEPGATGVGGPPGPAGEEGK--RG 291
QY 214 SGAQACESNL-----QGGNGGGGVAFYNLGMNMNMFQSGDIYMGSGSG--GGGGGAT 266
Db 292 ARGEPGSLPGPPGGERGGPGSRG-----PPGAD--GVAGPKGPGSGERGAP 335
QY 267 RPA 269
Db 336 GPA 338

Best local similarity 22.3%, Freq: NO: 0.22,
Matches 55: Conservative 15: Mismatches 54: Indels 116: Gaps 11:

Db 831 VLOPHLPKVDYTAQINSRLARLRLSLG-----DPKASTLPVRREQQQSSLLHPPPE 885
QY 110 AGGVVIT-----LHGREDIILSLTGTALPP----- 133
Db 886 KSPGEYVNIIEFGQGYLAGPATSRSPVRCPPQLHPAPREETGSEYEMNDLGPGR 945
QY 134 -----PAPCAGL---TYYLAGGQGVV---GNVAGSLIASGPPVLLMA 172
Db 946 ATWQEGGVVGLRIGPAGSATVCRTRSPVNSRDYMTMOIGCPRQSYVDTSPVA--P 1003
QY 173 ASFANAVYDLRPIEBEETPPRTTG-VQQQPEASQSVSEVTGSGA---QACESNLQGG-N 227
Db 1004 VSYADM---RTGIAAEKASLPRGTGAAPPSTASSASVYTPQATAEATHTSSLLGGPQ 1060
QY 228 GGGGV-AFYNLGMNMNFQ-----FSGGDIYGMS-----GG 257
Db 1061 GPGGMSAFTRVNLSPHNQSAKVIKIRADTQGCRRRHSSSETFSAPTRAGNTVPFGAGAAVGG 1120
QY 258 SGGGGGGAT 266
Db 1121 SGGGGGGGS 1129

RESULT 4

US-09-833-351-2
; Sequence 2, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; TITLE OF INVENTION: TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UTA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-833-351-2

Query Match 8.2%; Score 113.5; DB 3; Length 553;
Best Local Similarity 23.3%; Pred. No. 0.064;
Matches ,67; Conservative 26; Mismatches 104; Indels 91; Gaps 12;

QY 33 GRRPRGPA---GSKNKPPTTIIRD-----SPNVLRSHVLEVTSGSDISEAVS 79
Db 198 GROPFPAPPEQADGNAPGPQPPVRIQDIKTENGTCPPQPLSPAALGSGS--AAVP 255
QY 80 TYATRRGCGVCIISTGAVTWITQPAAPAGGGVITLHGRFDILSLTG--TALPPPPAPP 137
Db 256 KIESPDSSSSSLSGSPGSLPSARP-----LSLDGADSAPPPAPS 298
QY 138 G-----AGLTV-----YLAGGQGVVGNVAGSLIASGPPVVLMAASTANAVYDRLPIEEE 188
Db 299 APPPHSGGFGVDNIMTSLRSPQSAEAELSSGLLASA-----AASSRAGI----- 344
QY 189 ETTPPTTGVQO-----QQPEASQSVSEVTGSGAQACESNLQGGGGGVAFYNL-QNMN 241
Db 345 --APPLALGAYSPGSSLYSSPCSTSSAGSGGGGAGAGGAGGAGTGHCLNQLAMSL 402
QY 242 NNFQSGGDIYGMSSGGG-----GGGG 264
Db 403 YAAGERGGHLOGAFGAGGSAVDNPLPDYSLPPVTSSSSSLSHGGGG 450

RESULT 5

US-09-833-352-2
; Sequence 2, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UTA-029.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-833-352-2

Query Match 8.2%; Score 113.5; DB 4; Length 553;
Best Local Similarity 23.3%; Pred. No. 0.064;
Matches 67; Conservative 26; Mismatches 104; Indels 91; Gaps 12;

QY 33 GRRPRGPA---GSKNKPPTTIIRD-----SPNVLRSHVLEVTSGSDISEAVS 79
Db 198 GROPFPAPPEQADGNAPGPQPPVRIQDIKTENGTCPPQPLSPAALGSGS--AAVP 255
QY 80 TYATRRGCGVCIISTGAVTWITQPAAPAGGGVITLHGRFDILSLTG--TALPPPPAPP 137
Db 256 KIESPDSSSSSLSGSPGSLPSARP-----LSLDGADSAPPPAPS 298
QY 138 G-----AGLTV-----YLAGGQGVVGNVAGSLIASGPPVVLMAASTANAVYDRLPIEEE 188
Db 299 APPPHSGGFGVDNIMTSLRSPQSAEAELSSGLLASA-----AASSRAGI----- 344
QY 189 ETTPPTTGVQO-----QQPEASQSVSEVTGSGAQACESNLQGGGGGVAFYNL-QNMN 241
Db 345 --APPLALGAYSPGSSLYSSPCSTSSAGSGGGGAGAGGAGGAGTGHCLNQLAMSL 402
QY 242 NNFQSGGDIYGMSSGGG-----GGGG 264
Db 403 YAAGERGGHLOGAFGAGGSAVDNPLPDYSLPPVTSSSSSLSHGGGG 450

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	118.5	8.5	918	4	US-09-041-886-11	Sequence 11, Appl
2	116.5	8.4	1234	2	US-08-317-310A-15	Sequence 15, Appl
3	116.5	8.4	1234	5	PCR-US95-13041-15	Sequence 2, Appl
4	113.5	8.2	553	3	US-09-083-351-2	Sequence 2, Appl
5	113.5	8.2	553	4	US-09-083-352-2	Sequence 3, Appl
6	112	8.0	738	3	US-08-864-038A-3	Sequence 2, Appl
7	110	7.9	832	1	US-08-209-747-2	Sequence 2, Appl
8	110	7.9	832	1	US-08-458-298-2	Sequence 2, Appl
9	110	7.7	907	2	US-09-010-928B-4	Sequence 4, Appl
10	106	7.6	460	4	US-09-056-556-184	Sequence 184, Appl
11	104	7.5	1958	1	US-07-945-283-2	Sequence 2, Appl
12	103	7.4	760	1	US-08-195-152-2	Sequence 2, Appl
13	101.5	7.3	595	4	US-09-219-849-48	Sequence 48, Appl
14	101.5	7.3	595	4	US-09-219-849-50	Sequence 50, Appl
15	101.5	7.3	822	4	US-09-219-849-49	Sequence 49, Appl
16	101	7.3	1185	4	US-09-041-886-23	Sequence 23, Appl
17	100.5	7.2	412	4	US-09-027-064-4	Sequence 4, Appl
18	100.5	7.2	412	4	US-09-271-815-4	Sequence 4, Appl
19	99.5	7.1	378	2	US-08-986-217-6	Sequence 6, Appl
20	99	7.1	485	2	US-08-749-391-2	Sequence 2, Appl
21	99	7.1	485	4	US-09-390-200-2	Sequence 2, Appl
22	98	7.0	318	4	US-09-060-756-727	Sequence 727, Appl
23	98	7.0	334	4	US-09-060-756-728	Sequence 728, Appl
24	98	7.0	2055	1	US-08-175-155-46	Sequence 46, Appl
25	98	7.0	2055	1	US-08-477-509B-81	Sequence 81, Appl
26	98	7.0	2055	1	US-08-707-237A-52	Sequence 52, Appl
27	98	7.0	2055	3	US-08-482-085B-81	Sequence 81, Appl

Thu Jul 11 11:01:02 2002

[illegible]

Search completed: July 10, 2002, 22:20:18
Job time: 3611 sec

XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
PF
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 18-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 23-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150560.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134319.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149932.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0158458.
PR 29-SEP-1999; 99US-0158596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

```

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150360.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156590.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160769.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160988.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 572; DB 21; Length 272;

Best Local Similarity 47.8%; Pred. No. 1.8e-38;
Matches 120; Conservative 44; Mismatches 69; Indels 18; Gaps 6;
QY 5 RSEADEKAEITTTGGATSSATASGSSGRRPRGRPRGAGSKNRPKPTIITRDSPPNVLRS 64
Db 25 qsdeeqnsrdeqp---avepgsgsg-stgrprgrppgsknkpksvvtvkespnsiqsh 80
QY 65 VLEVTSGDISSEAVSYATRRGGVCIIISGTGAVTNTIRQAPAAPAGGGVITLHGRFDIL 124
Db 81 vleiatgadvaeslnafarrirgrgsvlsgglvtnvtlrqpa--asggvvalrggfell 138
QY 125 SLGTALPPP-APPAGAGLTVYLAGGGQVGVGNVAGSLIASGPPVLMASAFNAVYDRL 183
Db 139 smcgafliptsgspaaaagltilylagagqvgvggvggpliasgppviviaatfcnatyerl 198
QY 184 PIREETPPPTTGTGVOQQOPEASQSSEVTGSGGAQACESNLQGGGCGGVAFYNLGNM-- 241
Db 199 plieeq-----qqeqqlqliedkkkqkeendhesngnngmegsmgppmyumppnflp 249
QY 242 NNFQSGGGDIY 252
Db 250 nghqmaqhdvy 260

RESULT 11
AAG21496
ID AAG21496 standard; Protein; 310 AA.
XX AC AAG21496;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 24072.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 23-MAR-1999; 99US-0123548.
XX 25-MAR-1999; 99US-0125788.
XX 29-MAR-1999; 99US-0126264.
XX 01-APR-1999; 99US-0126785.
XX 06-APR-1999; 99US-0127462.
XX 08-APR-1999; 99US-0128234.
XX 16-APR-1999; 99US-0128714.
XX 19-APR-1999; 99US-0129845.
XX 21-APR-1999; 99US-0130077.
XX 23-APR-1999; 99US-0130449.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
```

```
Db      81 vleiatgadvaeslnafarrirgrgsvlsrgslvntvlrqpaa--asggvvslrgqfeill 138
Qy     125 SLTGALPPP-APPGAGGLTWYLAGGGGVVGNVGASLIASGPVVLMMAASFANAVYDRL 183
       l : | ||| :|||l||||lll||||lll||||lll:::|||l:l|l|
Db     139 smcgaflptsgpsaaaaglliylagaqqgvvgvagpliasgpviviaatfcnatyerl 198
Qy     184 PIEEEETPPRTTCGVOQQPEPASQSSEVTSGSAQAACESNLQGNGGGGVAFYNLGMNM-- 241
       l|||||   ||::|    :|||l|l|l|l|l|l|l|l|l|l|l|l|
Db     199 pieeeeq-----qqeqlqliedgkkkeendndnesngnneqsmqppmymnpnfip 249
Qy     242 NNQFSGGDIIY 252
       | | : |:|
Db     250 nghdmaghdy 260

RESULT_10
AAG54092 standard; Protein; 272 AA.
XX AC AAG54092;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 68934.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
```

Query Match	41.1%	Score 572:	DB 21:	Length 272:
-------------	-------	------------	--------	-------------

XX PF 14-NOV-2000; 2000WO-US31414.
 XX PR 17-NOV-1999; 99US-0166228.
 PR 17-APR-2000; 2000US-0197899.
 PR 22-AUG-2000; 2000US-0227439.
 XX (MEND-) MENDEL BIOTECHNOLOGY INC.
 PA (JIAN/) JIANG C.
 PA (HEAR/) HEARD J.
 PA (PINE/) PINEDA O.
 PA (PILG/) PILGRIM M.
 PA (ADAM/) ADAM L.
 PA (RIEC/) RIECHMANN J L.
 PA (YUGG/) YU G.
 PA (SAMA/) SAMAHA R.
 XX JIANG C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann JL;
 PI Yu G, Samaha R;
 XX WPI: 2001-335977/35.
 DR N-PSDB; AAD06672.
 XX Nucleic acids encoding plant transcription factor polypeptides, useful
 PT for altering the sugar sensing characteristics of plants and increasing
 PT yield, e.g. corn, potato and cotton plants -
 XX Claim 4; Page 144-145; 151pp; English.
 XX The patent relates to polynucleotides encoding 35 plant transcription
 CC factors which may be used to modify phenotype associated with a plant's
 CC sugar sensing characteristics and increasing yield when their expression
 CC level is altered. Sugars are central regulatory molecules that control
 CC aspects of physiology, metabolism and development. Therefore the control
 CC and proteins of the invention are useful for modifying the growth and
 CC germination rates of plants, photosynthesis, glyoxylate metabolism,
 CC respiration, starch and sucrose synthesis and degradation, pathogen
 CC response, wounding response, cell cycle regulation, pigmentation,
 CC flowering and senescence of plants and for modifying sink-source
 CC relationships in seeds, tubers, roots, and other storage organs leading
 CC to an increase in yield. The transcription factor polynucleotides and
 CC polypeptides may be used to alter the structure and developmental
 CC characteristics of plants such as soybean, wheat, corn, potato, cotton,
 CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,
 CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,
 CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,
 CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,
 CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.
 CC The present sequence is a homolog of Arabidopsis thaliana transcription
 CC factor.
 XX SQ Sequence 339 AA:
 Query Match 44.9%; Score 625.5; DB 22; Length 339;
 Best Local Similarity 49.2%; Pred. No. 1.1e-42;
 Matches 129; Conservative 28; Mismatches 48; Indels 57; Gaps 3;
 QY 4 NRSEADEAKETPTGGATSSATAGSSSSGRRPRGRPRAGSNKPKPTITRDPNVLR 63
 Db ntnsgsekgemslhgeggsgggsggeqmrtrprgrpgsknkpakpiliitrdsanalrt 147
 QY 64 HVLEVTGSDISEAVSYATRRGGVCITISGTGAVTNVTIROPAPAGGVITLHGRFDI 123
 Db hmeigdcidvdmatafarrgrgvcmsgtgsvtnvtrqpspp-gsvvslhgrfel 206
 QY 124 LSLTCTALPPAPPAGGLTYVLGGQGVVGGNVAGSLIASGPVILMAASFANAVYDRL 183
 Db lslsgsfllpppapaatglsylvlaggggvgvsgvpllcsgpvvvmaasfsnaayelr 266
 QY 184 PIERETPTPTRTTGVOQQPEASOSSEVTGSGACACESNLGGNGGGGVAFYNLGMWNN 243
 Db pteedemqtp-----vqggggggg----- 285

QY 244 FQFSGGDIYGMGGGGGGGA 265
 Db 286 -----ggggmgs 292
 RESULT 9
 AAG21497
 ID AAG21497 standard; Protein; 272 AA.
 XX AC AAG21497;
 XX DT 17-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 24073.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 XX PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.

PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149729.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156536.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157153.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 PR 29-OCT-1999; 99US-0162143.

Query Match 45.8%; Score 637.5; DB 21; Length 315;

Best Local Similarity 50.0%; Pred. No. 1.le-43;

Matches 132; Conservative 37; Mismatches 58; Indels 37; Gaps 5;

QY 17 PTGGATSSATAGSSSSRRPRGRPAGSKNKPPTIITRDSNVLRSHVLEVTSGDISIE 76
 Db 70 pregaveapt-----rrgrpagsknkpkpplfvtrdspnalkshvmetasgtdvie 122
 QY 77 AVSTYATRRGCGVCIISGTGAVTNVTIQP-----AAPAGGGVITLHGRFDILSLTATA 130
 Db 123 tlatfarrrrqgicllsgngtvanvltlqpstaavaaagggaavialggrfellsitgsf 182
 QY 131 LPPAPPAGGLTVYLAGCGQGVGNVAGSLIAGSPVVLMAASANAVYDLRPIEEBET 190
 Db 183 lpgpappggtgtitylaggqgqvvgvvpImaagpvmllaatfsnatyerlpleeeea 242
 QY 191 PP---PRTTGVOQQQPEASQSSEVTGSGAQACESNLOGNGGGGVAFVNLGMNNNFQF 246
 Db 243 aerqggggsggvgvvpq-----lqggggplsgaggggdnqqlpynmpgnlvs--- 290
 QY 247 SGDIYGMSSGGSGGGGATRPAPF 270
 Db 291 -----ngsggggggmsgqgeay 306

RESULT 8

AAE02571

ID AAE02571 standard; Protein; 339 AA.

XX

AC AAE02571;

XX

DT 10-AUG-2001 (first entry)

XX

DE A. thaliana transcription factor Gl068 homolog, G2657.

XX

KW Plant transcription factor; phenotype; sugar sensing characteristic;
 KW transgenic plant; plant yield; growth; germination; photosynthesis;
 KW glyoxylate metabolism; respiration; pathogen response; wounding response;
 KW cell cycle regulation; pigmentation; flowering; senescence; physiology;
 KW storage organ; metabolism.

XX

OS Arabidopsis thaliana.

XX

Key Location/Qualifiers

Domain Il6..l29

FT /note= "Conserved domain"

XX

PN WO200135725-A1.

XX

PR 25-MAY-2001.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 19:01:52 : Search time 1922.93 Seconds
(without alignments)
10599.682 Million cell updates/sec

Title: US-09-823-676-1

Perfect score: 974

Sequence: 1 cccccgacctgcctctaca.....cgaaatgcgaattagggtt 974

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hlg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

C	1	952.8	97.8	86710	8	ATF23E12	AL022604	Arabidops
	2	952.8	97.8	197859	8	ATFCHIRB3	AL161587	Arabidops
C	3	952.4	21.8	1323	8	AF194974	AF194974	Arabidops
	4	212.4	21.8	103192	8	AC007369	AC007369	Arabidops
C	5	208.6	21.4	88411	8	AC004667	AC004667	Arabidops
	6	205.2	21.1	122871	8	AC002387	AC002387	Arabidops
C	7	201.8	20.7	100806	8	AC015450	AC015450	Arabidops
	8	196.8	20.2	107600	8	ATF16J13	AL015450	Arabidops
C	9	196.8	20.2	190026	8	ATFCHIRV3	AL049638	Arabidops
C	10	191.6	19.7	197976	8	ATFCHIRV33	AL161533	Arabidops
	11	191.6	19.7	206606	8	ATFCA1	AL161539	Arabidops
C	12	190.4	19.5	1161	8	AF446359	297336	Arabidopsis
	13	190.4	19.5	81835	8	ATFCA9	AF446359	Arabidops
C	14	190.4	19.5	198067	8	ATFCHIRV4	297344	Arabidopsais
C	15	186.4	19.1	89350	8	ATT12H17	AL161547	Arabidops
	16	186.4	19.1	193550	8	ATFCHIRV58	AL021835	Arabidops
C	17	183	18.8	129845	2	AP004165	AL161558	Arabidops
	18	180.2	18.5	177418	2	OSJN00182	AL004165	Oryza sat
C	19	179.4	18.4	107865	8	ATT4C21	AL662981	Oryza sat
	20	178.8	18.4	103240	8	ATT32E16	AL162295	Arabidops
C	21	177.2	18.2	95310	8	NTAC011437	AL132975	Arabidops
	22	174	17.9	160363	2	AP003526	AL011437	Arabidops
C	23	170.4	17.5	116205	8	AC006931	AP003526	Oryza sat
C	24	166	17.0	1056834	2	AP004757	AC006931	Arabidops
	25	161.8	16.6	134712	2	AP003891	AP004757	Oryza sat
C	26	161.8	16.6	137569	2	AP004587	AP003891	Oryza sat
	27	160.8	16.5	108763	2	AP004020	AP004587	Oryza sat
C	28	148.8	15.3	147666	2	AP004635	AP004020	Oryza sat
	29	148.8	15.3	152381	2	AP004698	AP004635	Oryza sat
C	30	127.6	13.1	12319	6	B64635	AP004698	Oryza sat
	31	127.6	13.1	71807	8	AB025513	E64635 Method for	Arabidops
C	32	116.8	12.0	145847	2	AP004319	AB025513	Arabidops
	33	111.6	11.5	140466	8	AP003683	AP004319	Oryza sat
C	34	108.2	11.1	111686	8	F14L17	AP003683	Oryza sat
	35	101.2	10.4	279	8	AF184830	AC012188	Sequence
C	36	89.2	9.2	140327	2	AP003938	AF184830	Arabidops
C	37	87.8	9.0	169656	2	AP004680	AP003938	Oryza sat
	38	69.8	7.2	167764	2	AP004303	AP004680	Oryza sat
C	39	50.8	5.2	125020	9	AF429315	AP004303	Oryza sat
	40	45.2	4.6	590	8	HVU234403	AF429315	Homo sapi
C	41	45.2	4.6	124050	8	AF474373	AF429315	Homo sapi
	42	44.6	4.6	7218	6	I66494	AF474373	Hordeum v
C	43	43.8	4.5	125020	9	AF429315	I66494	Sequence 14
	44	42.2	4.3	1137	8	AMA132349	AF429315	Homo sapi
C	45	42	4.3	1694	8	ATAJ4119	AMA132349	Antirrhin
							AJ224119	Arabidops

ALIGNMENTS

RESULT 1

ATF23E12/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

schuelle@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES

source

1. .86710
 /organism="Arabidopsis thaliana"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

173. .1890

/gene="F23E12.10"

173. .374

/gene="F23E12.10"

/number=1

join(<173. .374,456. .919,1089. .1610,1678. .1890)

/gene="F23E12.10"

/codon_start=1

/product="hypothetical protein (fragment)"

/protein_id="CAA18726.1"

/db_xref="GI:3080407"

/db_xref="SPTREMBL:O65486"

/translation="FNAVAGCFEAVESVLWPSSSTSLPNTTSMVILSAVTASV
 VSEGLSEPAFKVPDVRSPGELPLYLILGALCVLSALSRCTSSMTSAVDLSNK
 DAGLPKAVFPMVGLSVGIIALVPIVEVLYWGFQNVLDLEKRPVKGLSADLLLOLVA
 VKIATAWCRASGLVGGYAPSLFTGGAGMAYKFIGLAQAQPDFNLISLEVASPO
 AYLGVNATAGVQVPLTALLFLFTQDYRIVPLLGAVMSSWITSGOSKRQET
 RETKTRKRQEAQVSLTSSDSESNLVCESLCLDDSLNQSELPKSFVSEA
 MKTRFATVMSTSLSEALTRMLIEKQSCALIVDPNIFLITLSDIQEFSGARKEG
 NRPKDIFVNDICSRSGKCKVPWTVPMDLLAAQTINMKHLSHVAVVSGSIDAPRI
 HPGVLDRECIITLPR"

375. .455

/gene="F23E12.10"

/number=1

456. .919

/gene="F23E12.10"

/number=2

920. .1088

/gene="F23E12.10"

/number=2

1089. .1610

/gene="F23E12.10"

/number=3

1611. .1677

/gene="F23E12.10"

/number=3

1678. .1890

/gene="F23E12.10"

/number=4

3124. .3241

/gene="F23E12.20"

/number=1

3124. .4428

/gene="F23E12.20"

join(3124. .3241,3323. .3465,3595. .3696,3945. .4107,

4211. .4428)

/gene="F23E12.20"

/note="similarity to dihydroflavonol-4-reductase, Gerbera

x sp., PIR2:S35189"

/codon_start=1

/product="putative protein"

/protein_id="CAA18727.1"

/db_xref="GI:3080408"

/db_xref="SPTREMBL:O65487"

/translation="MDQAKGVCTVTCASGFLASWLKRLLEGEYVIGTVRDPGNEKK
 LAHLWKLEAKERLRLVKADLMESGFDAIMGCQGVFTASPEILRPALEGTANVLR
 SCRKNPSLRKRVLTSSSTVWYALSKTLAEQAQWKFSEENGIDLVTLVLSFLVGSPLP
 PDLCTASDLVLLGKETGEKQFMQGMGYDDVARTHVVFHEAAQGRYICSSNV
 ISLELVSPLSARYPSLPIPKRFDTPLN"

3242. .3322

intron)

/gene="F23E12.20"
 /number=1
 3323. .3465
 /gene="F23E12.20"
 /number=2
 3466. .3594
 /gene="F23E12.20"
 /number=2
 3595. .3696
 /gene="F23E12.20"
 /number=3
 3697. .3944
 /gene="F23E12.20"
 /number=3
 3945. .4107
 /gene="F23E12.20"
 /number=4
 4108. .4210
 /gene="F23E12.20"
 /number=4
 4211. .4428
 /gene="F23E12.20"
 /number=5
 complement(join(4939. .5013,5109. .5195,5307. .5342,
 5451. .5651,5725. .5811,6161. .6163))
 /gene="F23E12.30"
 4939. .6163
 /gene="F23E12.30"
 complement(join(4939. .5013,5109. .5195,5307. .5342,
 5451. .5651,5725. .5811,6161. .6163))
 /gene="F23E12.30"
 /note="Contains Clathrin adaptor complexes small chain
 signature [VVKRYASLYP]
 contains EST gb:T44219"
 /codon_start=1
 /product="clathrin assembly protein AP19 homolog"
 /protein_id="CAA18728.1"
 /db_xref="GI:3080409"
 /db_xref="SPTREMBL:O23685"
 /translation="MIHEVLVSRQGVKRLTKWYSPYAKRSKRVIRELSGVILNRGP
 KLCNFEWVRGKVKYKRYKASLYFCIDQEDNEVLEIHHYVELDRFSGVCELD
 LIFNFHKAYYILDELLIAGLOESSKKTARIISQDQLVEVAKEEASSISIIAQT
 NR"
 complement(4939. .5013)
 /gene="F23E12.30"
 /number=1
 complement(5014. .5108)
 /gene="F23E12.30"
 /number=1
 complement(5109. .5195)
 /gene="F23E12.30"
 /number=2
 complement(5196. .5306)
 /gene="F23E12.30"
 /number=2
 complement(5307. .5342)
 /gene="F23E12.30"
 /number=3
 complement(5343. .5450)
 /gene="F23E12.30"
 /number=3
 complement(5451. .5651)
 /gene="F23E12.30"
 /number=4
 complement(5652. .5724)
 /gene="F23E12.30"
 /number=4
 complement(5725. .5811)
 /gene="F23E12.30"
 /number=5
 complement(5812. .6160)
 /gene="F23E12.30"
 /number=5

Qy	547	aattgcttcggagccggttagttgatggctcttcttttgcaaacgcagtttatgatag	608
Db	8560	AATTGCTTCGGGACCGGTAGTGTATGGCTGCTTTTGCAAAGCAGCTTTATGATAG	8501
Qy	607	gttacggattgaagaagaaccccccccgaggagaccaccccgggtgcagcagcagca	666
Db	8500	GTTACCAGATTGAAGAGAGAAGAACCCGCCCGCGAGAACCCACCGGGGTGCACGACGASCA	8441
Qy	667	gccggaggcgcttcagtcgaggttacggggagtggggcccgagcgctgtgagtcacaa	726
Db	8440	GCCGGAGGCGCTTCAGTCTCGAGGTTACGGGGAGTGGGCCCGAGCGCTGTGAGTCAA	8381
Qy	727	cctccaaagtgcgaatggttgaggagggttgtctcttcataaatcttgggaatgaatatgaa	786
Db	8380	CTTCCAAGGTGGAAATGTTGGAGGAGGTGTGCTTTCTACAAATCTTGGAAATGAATATGAA	8321
Qy	787	caattttcaattctccggggagatattacqglatgagcgcggtagcgaggagagctg	846
Db	8320	CAATTTTCAATTCTCGGGSGAGATATTACGGTATGAGCGCGGTACCGGAGGAGGTGG	8361
Qy	847	tggcgtgcgactagaccgcgcttttagagtttttagogtttggdgacaccttttgttc	906
Db	8260	TGCGGTGGCACTAGACCCGCGTTTTTAGAGTTTTAGCGTTTTTGGTGACACTTTTGTTC	8201
Qy	907	gttgcgtgtttgacctcaaacctactagctactagctatagctggttcgcaaatgcgaat	966
Db	8200	GTTTGGCTGTTTGACCTCAAACTACTAGCTACTAGCTATAGCGGTTGCGAAAATGCGAAT	8141
Qy	967	attaggtt 974	
Db	8140	ATTAGGTT 8133	
<hr/>			
RESULT	2		
LOCUS	ATCHRIV83		
DEFINITION	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 83.		
ACCESSION	AL161587		
VERSION	AL161587.2	GI:7270470	
KEYWORDS			
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophytas; Spermatopsida; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis..		
JOURNAL	1 (bases 16641 to 103350)		
REFERENCE	Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A., Mewes,H.W., Lemcke,K. and Mayer,K.F.X. Unpublished		
AUTHORS	2 (bases 98352 to 125759) Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and Mayer,K.F.X. Unpublished		
JOURNAL	3 (bases 120761 to 197859)		
REFERENCE	Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and Mayer,K.F.X. Unpublished		
AUTHORS	4 (bases 1 to 24256) Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and Mayer,K.F.X. Unpublished		
JOURNAL	5 (bases 1 to 197859)		
REFERENCE	EU Arabidopsis sequencing, project.		
AUTHORS	Submitted (10-NAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk		
JOURNAL	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			

Qy	787	caattttcaattcccg9999gaataatttaccggtatgcg9cg9tagcgaggaggtgg	846
Db	95031	CAATTTTCAATTCCTCGGGGGAGATATTTACGGTATGACGCGGTAGCGGAGGAGTGG	95090
Qy	847	tggcgggtgcgactagaccgcgcttttagagtttttagcgcttttggcgacaccttttggtc	906
Db	95091	TGCGGTGCGACTAGACCCGCTTTTAGAGTTTTAGCGTTTTGGTGACACCTTTTGTTGC	95150
Qy	907	gtttgcggtttgacctcaaacactagctactagctatgcg9gttgcgaatgcgaat	966
Db	95151	GTTTGGCGTGTTCACCTCAACTACTAGCTACTAGCTATAGCGTTCGGAATCGCAAT	95210
Qy	967	attaggtt 974	
Db	95211	ATTAGGTT 95218	
RESULT	3		
AF194974		AF194974	
LOCUS		Arabidopsis thaliana ESCAROLA (ESC) mRNA	linear
DEFINITION		1323 bp	PLN 10-NOV-1999
ACCESSION		Arabidopsis thaliana ESCAROLA (ESC) mRNA, complete cds.	
VERSION		AF194974	
KEYWORDS		AF194974.1	GI:6319179
SOURCE			
ORGANISM		thale cress.	
REFERENCE			
AUTHORS		Arabidopsis thaliana	
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.	
		1. (bases 1 to 1323)	
		Weigel, D., Ann, J.H., Blazquez, M.A., Borevitz, J., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z., Xia, Y., Dixon, R.A., Harrison, M.J., Lamb, C.J., Vancovsky, M.F. and Chory, J.	
TITLE		Activation Tagging in Arabidopsis	
JOURNAL		Unpublished	
REFERENCE		2. (bases 1 to 1323)	
AUTHORS		Nguyen, J.T., Christensen, S.K. and Weigel, D.	
TITLE		Direct Submission	
JOURNAL		Submitted (13-OCT-1999) Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA	
FEATURES			
SOURCE		Location/Qualifiers	
		1..1323	
		/organism="Arabidopsis thaliana"	
		/db_xref="taxon:3702"	
5'UTR		1..274	
gene		/gene="ESC"	
		1..1323	
		/gene="ESC"	
CDS		275..1210	
		/gene="ESC"	
		/note="overexpression causes late flowering, modified leaf development; similar to F9H16.12 encoded by GenBank Accession Number AC007369; similar to AT-hook proteins and pea PDI homolog"	
		/codon_start=1	
		/product="ESCAROLA"	
		/protein_id="AAF0197.1"	
		/db_xref="GI:6319180"	
		/translation="MEGGYEQGGASRYFNLFRPEIHHQQLPOGGINLIDQHOOHQIQHQOQPSDDSRSHDNKHQODSDPNTSSAPGKRPRPGSGKNKAPKPIIVTSDPSNALSHLEVSPGADIVESVSYARRRGVSVLGGNGTVSVNLTLPQPTVPGNGGVSGGGVVTLHGRRFELSLTGTLPAPPAGGAGSIFLAGCGQGVSGSVAPLIAASQVILMAASFNAYFRELPIEEEEEGGGGGGGGGPQMQAAPSAAPPSPVTCGQPLGVNGVGGYFGSGDPHLLGWAGTGPSRPF"	
3'UTR		1211..1323	
		/gene="ESC"	
BASE COUNT		317 a 286 c 346 g 374 t	
ORIGIN			

Query Match	21.8%;	Score 212.4;	DB 8;	Length 1323;
Best Local Similarity	62.8%;	Pred. No. 5.6e-49;		
Matches 358;	Conservative 0;	Mismatches 191;	Indels 21;	Gaps 1;
QY 80	gaagcagacgaagcaaaagccgagaccactccacccggtgagagcaccagctcagccaca	139		
DB 449	GAATCTGACCATTAACAAGATCATCATCAACAGGGTCGACCGGATTCAGACCCGAAT	508		
QY 140	gctctggtctctcccgagcgtgctccacgtggtgctgctcgcagggtcccaaaacaaa	199		
DB 509	ACATCAAGCTCAGCAGCGGAAAGCTCCACGTGGACGCTCCACAGGATCTAAGAACAA	568		
QY 200	cccaaacctccagacattataactagaatctcctaacgtctcttagatcacagcttctt	259		
DB 569	GCAAGCCACCATCATAGTAAGTCTGATGATCCCAACCGCCTTAGATCTCAGCTTCT	628		
QY 260	gaagtcacctccggttcgacatataccagggcaggtctccacactcgcactcgcgcgc	319		
DB 629	GAAGTATCTCTGGAGCTGACATAGTTGAGAGTGTTCACGCTACGCTAGGAGGAGGG	688		
QY 320	tgcgcggttgattataagcgcagcgggtgcgtcacttaacgtcacgtacgcgaacct	379		
DB 689	AGAGCGCTCCGCTTTTAGGAGGAACAGCACCGTATCTAACGTCACTCTCCGTACCCA	748		
QY 380	gcgggtcc-----cggtcgtggtgaggtgtgattaccctgcagtgt	418		
DB 749	GTCACCTCTGGAATGGCGGTGGTGTCTCCGGAGGAGGAGGAGTGTGTACTTTACATGA	808		
QY 419	cggttgacatttctcttgaccggttactgcgtctccaccgctcaccacccggagca	478		
DB 809	AGGTTTGAATCTTTCGCTAACGGGACTGTTTTCGCCACCTCTCCACCGCTGTGGCC	868		
QY 479	gaaggttcagcgtgatctagcgcaggtgcaaggcaagttgttagaggaatgtggct	538		
DB 869	GGTGGTTGTCTATATTTTACCGGAGGCAAGGTCAGTGTGTCGAGAGGACGCTGTG	928		
QY 539	ggtctgtaattgcttcggagcgcgttagtgtgtagtgcgtctcttcttcaaacgcagtt	598		
DB 929	GCTCCCTTATTCATCAGCTCAGCTCGGTATATACTAATGGCGCTTCGTCTCAATGCGGT	988		
QY 599	tatgataggtaccgattcaaggaagaa	628		
DB 989	TTGAGAGACTACCGATTGAGGAGGAGAA	1018		
RESULT 4				
AC007369/c				
LOCUS	103192 bp	DNA	linear	PLN 17-MAY-1999
DEFINITION	Arabidopsis thaliana chromosome I BAC F9H16 genomic sequence, complete sequence.			
ACCESSION	AC007369			
VERSION	AC007369.2			
KEYWORDS	HTG.			
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.			
AUTHORS	1 (bases 1 to 103192)			
	Fedorov, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altai, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.			
	Unpublished			
JOURNAL	2 (bases 1 to 103192)			
REFERENCE	Fedorov, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altai, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.			
AUTHORS	Submitted (24-APR-1999) DNA Sequencing and Technology Center,			
	Direct Submission			
TITLE	Arabidopsis thaliana			
JOURNAL	Submitted (24-APR-1999) DNA Sequencing and Technology Center,			

Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA	3 (bases 1 to 103192)
Fedorov, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altai, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.	4 (bases 1 to 103192)
Direct Submission	4 (bases 1 to 103192)
Submitted (07-MAY-1999) DNA Sequencing and Technology Center,	4 (bases 1 to 103192)
Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA	4 (bases 1 to 103192)
Fedorov, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altai, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.	4 (bases 1 to 103192)
Direct Submission	4 (bases 1 to 103192)
Submitted (17-MAY-1999) DNA Sequencing and Technology Center,	4 (bases 1 to 103192)
Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA	4 (bases 1 to 103192)
On May 7, 1999 this sequence version replaced gi:4678189.	4 (bases 1 to 103192)
e-mail for correspondence: arab@sequence.stanford.edu	4 (bases 1 to 103192)
Genes with similarity to proteins in the databases are described as 'putative', 'like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'.	4 (bases 1 to 103192)
The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge, http://genomic.stanford.edu/~chris/GENSCANW.html), Fexa (V.Solovyev & A.Salamov, Sanger Centre, http://genomic.sanger.ac.uk/), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).	4 (bases 1 to 103192)
Location/Qualifiers	4 (bases 1 to 103192)
1. 103192	4 (bases 1 to 103192)
/organism="Arabidopsis thaliana"	4 (bases 1 to 103192)
/cultivar="Columbia"	4 (bases 1 to 103192)
/db_xref="taxon:3702"	4 (bases 1 to 103192)
/chromosome="I"	4 (bases 1 to 103192)
/map="mi203"	4 (bases 1 to 103192)
/clone="F9H16"	4 (bases 1 to 103192)
/complement(8784..9859)	4 (bases 1 to 103192)
/gene="F9H16.1"	4 (bases 1 to 103192)
/complement(join(8784..9101,9359..9574,9662..9859))	4 (bases 1 to 103192)
/note="Unknown protein; Location of ESTs 152C14T7, gb T88012, 173P7T7, gb H36502 and 107E18T7, gb T22775"	4 (bases 1 to 103192)
gb T88012, 173P7T7, gb H36502 and 107E18T7, gb T22775"	4 (bases 1 to 103192)
/codon_start=1	4 (bases 1 to 103192)
/protein_id="AAD30591.1"	4 (bases 1 to 103192)
/db_xref="GI:4836888"	4 (bases 1 to 103192)
/translation="MRTKEEDYTPPLMPMLRGSYFVPCSIHVDNKNNECNFLC	4 (bases 1 to 103192)
DCAGNAFCYCLVKKHRRVQIRRSYHVRVNEIQFIDACVQTYIINSAKIVE	4 (bases 1 to 103192)
YNERPQRIKGVNTCEICRSLDLSFRCSGLGKGGMRGDLSTLSLAKGHRRE	4 (bases 1 to 103192)
YLGGSDSEATPTTKMRKTNFNLMSGLSISTVRFDYGPNGDQRSSSGDEGGFSF	4 (bases 1 to 103192)
SPTPTPIYHNRSSRRKGVPHRAPF"	4 (bases 1 to 103192)
15929..17348	4 (bases 1 to 103192)
/gene="F9H16.2"	4 (bases 1 to 103192)
/join(15929..16340,16804..17348)	4 (bases 1 to 103192)
/gene="F9H16.2"	4 (bases 1 to 103192)
/note="Hypothetical protein"	4 (bases 1 to 103192)
/codon_start=1	4 (bases 1 to 103192)
/protein_id="AAD30592.1"	4 (bases 1 to 103192)
/db_xref="GI:4836889"	4 (bases 1 to 103192)
/translation="MRTKSIVINSKNDTSTISATLRRSASNTTPSAAYVTNDL	4 (bases 1 to 103192)
NKKPRSRVEFFAPQVQGEEMVHFHPOHVLVKVELPDIIYTCAGCKEGAGVRYVC	4 (bases 1 to 103192)
QECQYOLHEFCALAPQLKSPHPHYQLLFFAKPAKGIVKICDVCGRSPKGYTR	4 (bases 1 to 103192)
KACSFQMPGDCAMLSPLSSSLHHPLRLPLSSAGSTGGSGGLGCEGRKR	4 (bases 1 to 103192)
TGRVYRCTVDYHLHACVCAAAVNGLRANGHKGRKSPAVLGTAAARLASQVVIDFLG	4 (bases 1 to 103192)
GIIDLGEGVGEALIDGVTRCGGGGGGGGTVPVPRGG"	4 (bases 1 to 103192)
complement(18939..22830)	4 (bases 1 to 103192)

```

CDS
/gene="F9H16.3"
complement(join(18939..19356,19441..20040,20119..20263,
20346..20575..20692,20776..20883,20975..21049,
21141..21943,22041..22246,22334..22830))
/gene="F9H16.3"
/notes="Unknown protein; location of ESTs: 11IG20XP 3',
gb|AA585893..FA1314 5' end, gb|237590 and 11IG20T7,
gb|142259"
/codon_start=1
/protein_id="AAD30593.1"
/db_xref="GI:4836890"
/translation="MDEGCAVAAWPFTHQSIGKRDLYYPMSNRLVQSQRDEWN
SKWMDRRPEPAKVDVVEVEFDTLIRNKGEEGLDNLGSLTAVEETTTQNV
RNKVKYRSGPGNTIPCMQVQDCTEDLSHAKYHRRHKVCEVSHKATKALQKQMR
CQCGRFLLSDFEGKRCRRRRKTKTQPEEVASGVVPCGNHDTTNTAN
AMMDLALITALACQAGNAVKPVGSPAPVDRQQLILNKINALPLMDLVSKLNN
IGSLARKNNDHTVPQNDMGASPTMDLAVLSTTLGSSSPDALTLISOGFGND
SEKTKLSYEVNCTVNLBKRTFGSSVGGERSNSQSPDSRQDTRSSLSQL
FTSSPEDESPTVASSRKYYSASNNVEDRSPSSPYMQLFLPLQASPTMRKNNK
NSSPTGCLPLEFGASNRGAADPNFKFGQOQSGYASSGSDYSPPLNSDAQDRGTI
VFKLDKQPSQLPGLTRSEIYNLSNIPSEMSYIRPGCVVLSVIVAMSPAAWQLEQ
KLLOKGLLQNSPDSFWRNARFVNTGROLASHKNGKVRCSKWRWNSPELVSQ
VAVAGETSILVVRGSLTNDGISIRCTHMGSYNAMEVTRAVCRQTIFDELNVSEFV
QNVHPGFLGRCFIEVNGFRGDSRPLITANASICKELNLEGEHPKQSDWTEQAQS
NRRGTSREELVCLNELGWLFOKQNSLSELRQDSFLAREKFLVCSVERXDYLIR
TLIDMLVERNINDELNEALMLAEIQLLNARVKRSTKMWELLIIHYLVNPLTSS
RKFPPLNTPGGITPLHLAECTSGDDMTDLTNDPQETGLSSWNTLRDATQTPY
SYAALRNHNNTNSLVKRLADRNKQVSLNIEHEVVDQTLGSLKRLSLEMNKSSSCAS
CATVALKYQVRVSSQRLPPTPIIHSMLAVATVCVCVPMHAFPIVRQGSFHSWGL
DYGSI"
complement(27992..33534)
/gene="F9H16.4"
complement(join(27992..28280,28670..28985,29062..29743,
29822..29920,30018..30212,30431..30610,30713..30871,
30713..30871,30958..33534))
/gene="F9H16.4"
/notes="Hypothetical protein"
/codon_start=1
/protein_id="AAD30594.1"
/db_xref="GI:4836891"
/translation="MPVEVEHQGEVSVKVDENATEIKPEVYVVSATKEDVNGISIG
GSNNGNDVDSQDLEFTEINDVGDDEVDYKPVDDANVEDKLDKGENVYKWDAPSI
ADDVLGVSQDQSLFEKSELESTDDGVEVEIPEKSEVEDSLKSDVQHPGNGHLES
GLEGKSEVEQLHDESVGSKDLTKNNVEPEVEIETSESDTDECHQDGLTENQEK
SDRDLVSQDLKLNENKHPVDSEVSELYSAKYSPTEPSDGGNDLQPVITDPA
ELTNGSEVDHVCSEPVTVLEPVSVENGHPPESELSERSDVFTFVAEKVNASDGE
VLPDCTVDVYVSVSDVPAETALNAISLDSPQSKDQSVENGSKSESDSKMQS
EIGAVDQGSVSDSINTHPESQDASDPTCDGQKQHSSEVKEVLDAPASEEISDAVI
VAKDNGSEAAISDGLSTNQSGSEDSISGLVEKLPSHALHEVVSANDTSIVSDOT
KSOGLSDHGVDTNQTIDQDCSABELEVTVNVPKAPNEKVQDNGSEGNLVNGDVCL
NSABEAKELPTGLDSGNASHESAEITLSDNIDPLSLDLTKTAVSDFAESSVAGEID
AVAMESAASIKOCARAHVAPSTIEDGETDREVCNGSEVNVKTPPVAVREDIPPE
VSEMEESDVKERSINTDESVATASAEIKTQADLESKVTSTDIHFGAKDQVDS
OPARKNGKLNINEIRCLFSLVENQDQVDSIYKLLCSNVYDRTDDKDVASGEVSV
LDASEGTVAAEIEKRPFFELPVRPYDDDEKLAQKHAEEVDQTKQNRDLARADIQ
KIRACIKDYDISYKVAWEERSARKAMHSKROIEALQSMISRVKSAASVDDISRVNRAI
CKDYDISYKVAWEERSARKAMHSKROIEALQSMISRVKSAASVDDISRVNRAI
MQHTLSLNEKEGFMREIKOLKOLREIISMSGTKDVKQALDEKTEERLKYLRKE
LDALNRLSAAEITTKAAKKCDGEMPAQSKLQEPRAADVQAEAFVHLQDLKQQR
EKNYFFKYBNSRAASEMALKDRAALQSKLQDQVFNEMNNNDDEFRYYKVSNT
RSTFRRLGTLDRGSETPRITTYAPRTDKLRTSSDRAEKHEAPVPAQKEVYKEFG
SKVNNNEKQDTEQKQSTTKKVAQDPQPSIVTELVSKEIEKSTAKPEEPPEEP
KLTKEEBELIKKEERKQKAEAMKEQHLREIEIAKAEAMERKKREKAKARVLK
AQKEAREKVAKYFCHSVTQIGFRSVSLARTREEAKEEGKGFYIRDSNRKPNPS
DSRNSRNPKGDRNSKETNRGETNEEISQTEITVSETKQVKGIGSAFKPKRKQET
AAMDVWTHSCDHRHVPSSRSTIVRSCHISHPVPHQCQEIHLDFPSLFKDRAEK
HEVPPVPSALDSVLYLPLHQNLHLDQLLEQSMRRTTRTASRSSATFCLDNRSLSLTPP"
37958..45281
/gene="F9H16.5"
join(37958..40897,41604..41993,42096..45281)
/gene="F9H16.5"
/notes="Putative RNA helicase; Highly similar to human U5
snRNP-specific 200KD protein, gi|3255965 and yeast

pre-mRNA splicing helicase BRR2, gi|731380 and other RNA
helicases."
/codon_start=1
/protein_id="AAD30595.1"
/db_xref="GI:4836892"
/translation="MANLGGAHAHARFKOYEVANSLSVLTNDNPROTHPEPTGEPE
TLNGKIDPRSGDRVAKRQPELEDKLLKSKKKRQVDDVWVNIROSKRRRREESVY
TDYDAYQPKTETRAAYEPAMGLQKQKQGGPPSIVSAADEIILAVLKNDAFRPE
KMBEIEKLNKIEHFEFDQVLSIGKLTIDFQEGDSDGGRANDDEGDDLDGGLVAFVE
ENBDESDSPDDEDEDEDEPTRTGGMQVDAGINDEADAGDEGNLNVQDIDA
YWLQKTSIQAEQIDPQCCVLAELKILAGDDRVBDEKLLMHQYKESFLVFL
LNRLLKVVWCTRLARADQBERNIEEMRGLGPELTAIVEOLHATRAKAREENLQ
KLSNEARRLLDETGGDGRDVAADRSFQSVQGRQMLDLESIAEDQGGLLMAN
KKDLPGRSVRSHGKYDEYVHVPSKVKVDNRNKLKITEPMDWAPQAPRGMQOLNRV
QSKYDTALFAENILGCTAGKTNAMLTILQOELMNRNDGTNGYHGDYKRVVA
PMKALVAEVVNSLNKDYIVIRELSQSLTGREIETQIIVTPPKWDITTRKS
EDYVTLQVRLIIDEIHLHNRGPVLESIVARTLQIETTKENTRLVGLSATLPNY
EDVALFLRVDLKKGLFKFDRYRPPVPHQYIGISVKKPLQRFOLMNDLCYOKVLGA
GKQVLIFVHSRKETSKTARAIKDTAMANDTSLREFKEDSVDFVLSHEDIKNSDL
KDILPYGFATHHAGLSRGDEIYETLFSQGHVOVLYSTATLNGVNLPHARTVIRQTQ
VYNPEKGAHNELSPLDVMQMLGRAPQIDQHGEGIIITGISELYLSLMEQDPIE
SQFISKLADQLNAEIVLGTQVNAEACHMLGTIYIRVRNPTLYGLAPDALADKVY
LEPRADLIHSAATILDKNNLVYDRKSGYFQVTDLGRIASYIYTHGTIAYNHLK
PTMGDIDLVLRLFSDEFKVTYVRODEKMLAKLLDRVPIPIKETLEEPSAKINVLQ
AYISQLKEGLSITSDMWYITQSAGRLVRLVILVLRGWAQLAELAKALNSKMGKRM
WSYQTPLRPHGLSNDILQLEKDLWERYDYLSQELGELIRSPMKPGLPHKFIHQ
FPKVTLSAHVQPIRTVRLNVLNLPDFLWDEKIHKIVVEFFIIVEDNDGEKILHHEY
```

Query Match 21.8% Score 212.4; DB 8; Length 103192;

Best Local Similarity 62.8% Pred. No. 8.6e-49;

Matches 358; Conservative 0; Mismatches 191; Indels 21; Gaps 1;

QY 80 gaagcagacaaagccagagaccactccacccggtgagagccaccagctcagcaca 139

Db 74674 GAATCTGACCATTCAAACAAGATCATCATCAACAGGGTCGACCCCGAATTCAGACCCGAAT 74615

QY 140 gcctctgctcttctccgcagctcctccacgtgctcgtcgtcaggttccaggttcccaaaacaa 199

Db 74614 ACATCAAGCTCAGCACCGGAAACGTCACGTCGAGCTCCACAGGATCTAAGAACAA 74555

QY 200 cccaaacctccagcagattataactagatagtagtccctaaagctcctagatcagcagctctt 259

Db 74554 GCCAAGCCACCGATCATAGTAATCGTATAGCCCAACCGCTTAGATCTCAGCTCTT 74495

QY 260 gaagtcacctccgttcggacatatccgagcgagctccacccctacccactcgcctcgcgc 319

Db 74494 GAAGTATCTCTGGAGCTGACATAGTTGAGAGTGTTCACAGCTACGCTAGGAGGAGGG 74435

QY 320 tgcggcgtttgattataagcgcagcgggtgcggctcaactaaagctcagatcagcagcagcct 379

Db 74434 AGAGGCGCTCTCCGTTTGTAGGAGAAACGGCAGCTATCTAACGTCACCTCTCCGTCAGCCA 74375

QY 380 gcggcgc-----cggtcgtggagggtgattaccctgcaggt 418

Db 74374 GTCACTCTCGGAATGGCGGTGTGTGTCGGAGGAGGAGGAGTGTGACTTTACATGA 74315

QY 419 cggttgacattttgtctttgacccgtactgcgcttccaccgctcagcaccacccggagcga 478

Db 74314 AGGTTTGAGATCTTCTCGCTTAACGGGACTGTTTTCGCCACCTCTCTCCACCGCTGTCGC 74255

QY 479 ggaagttgacggtgtatctagccgaggtccaaggaagaagttgtaggaaggaagtgagct 538

Db 74254 GGTGGTTTGTCTATATATTTAGCCGAGGAGGCAAGGTGAGGTGTCGGAGGAAGCGTGTG 74195

QY 539 ggttcgtaattgcttcggaccggttagtggtagggctgctctcttttgcgaacagctt 598

Db 74194 GCTCCCTTATTCATCAGCTCCGCTTACTTAATGGCGGCTTCGTTCTCAATGCGGTT 74135

QY 599 tatgatagttaccgattgaagaggaaga 628

Db 74134 TTCGAGACTACCGATTGAGGAGGAGGAA 74105

RESULT 5	AC004667	88411 bp	DNA	linear	PLN 05-APR-2000
LOCUS	Arabidopsis thaliana chromosome II section 193 of 255 of the complete sequence. Sequence from clones T4C15, T32F12.				
DEFINITION	AC004667 AE002093				
VERSION	AC004667.2	GI:6598430			
KEYWORDS	HTG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Rhee, S., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J., Ronning, C.M., Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Niernman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.				
TITLE	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana				
JOURNAL	Nature 402 (6763), 761-768 (1999)				
MEDLINE	20083487				
PUBMED	10617197				
REFERENCE	2 (bases 1 to 88411)				
AUTHORS	Lin, X.				
TITLE	Direct submission				
JOURNAL	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA				
COMMENT	On Dec 17, 1999 this sequence version replaced gi:3668073. The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tadb/at/at.html).				
	Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (http://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNScan-SE/). Simple repeats were identified by RepeatMasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.				
	We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.				
	This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.				
	Address all correspondence to: at@tigr.org .				
FEATURES	source	Location/Qualifiers			
		1. .88411			
		/organism="Arabidopsis thaliana"			
		/cultivar="Columbia"			
		/db_xref="taxon:3702"			
		/chromosome="II"			
	misc_feature	complement(<1..88210)			
	repeat_region	/note="Sequence from clone T4C15"			
		2372..2441			
		/rpt_family="(TAAA)n"			
	mRNA	Join(<2768..2959,3319..3416,3488..3524,3636..3732,4160..4262,5931..6022,6392..6444,6497..6656,6766..7026,7524..7603,7800..>7997)			
		/gene="At2g35070"			
	gene	<2768..>7997			
		/gene="At2g35070"			
		/note="T4C15.26; predicted by genscan"			
		Join(2768..2959,3319..3416,3488..3524,3636..3732,4160..4262,5931..6022,6392..6444,6497..6656,6766..7026,7524..7603,7800..7997)			
		/gene="At2g35070"			
		/note="hypothetical protein"			
		/codon_start=1			
		/protein_id="AAC61828.1"			
		/db_xref="GI:3668096"			
		/translation="MAMKRTISIVIRIGLSYRSLSSGNHVDKLGVMKALEAINVPS TQAEALTGATSGFESVMGKEYKSTVAEEFKMRADIEKMRADIATIQAEERCDHDN HIHKVEQGNELKDEFDATKIKVIKYAITLIPVAAETAASCRVMTSFRNTNVH LGAVHMRQYQTCIDKMGINEFSFGRGSGFGEHNGESFPFPIWEYKQVAVVDLM SFVQREKSMVRACNAQGHAI SLDLETAALKMSADLNRTIEPSSVSKALLYESLPKS ISFLVKSSNLEGTLTDEAESVESVVMLLQDFDKLTKMETIILQLOTPNQE HCSSKWCNRVAVQCVDFSKQTIIRFGSMRNFDHVAIVSSLSFSSRVKEIRLQE HIDLLMLPAVFAVFAEVRSLSVVPEATQAHNRSSYTEVSLFELATSVKASPLILGRLL CKEHP"			
	mRNA	complement(Join(<8495..8749,9119..9154,9181..9248,9563..9584,10086..10390,10656..>10698))			
		/gene="At2g35080"			
	gene	complement(<8495..>10698)			
		/gene="At2g35080"			
		/note="T4C15.25; predicted by genscan"			
		complement(Join(8495..8749,9119..9154,9181..9248,9563..9584,10086..10390,10656..10698))			
		/gene="At2g35080"			
		/note="hypothetical protein"			
		/codon_start=1			
		/protein_id="AAC61827.1"			
		/db_xref="GI:3668095"			
		/translation="MVQANTYDARNELKEITDFLLDLTETADVILGYSLSKLGE TWQDHTFAFSHNDWISFNKEELKONTNKMWSEYVAKGIEDSQISDCGMVDDY LEDGFGSGNGRHFPKFIITFSRKISEKTSVDLNHPFSLPNIRTOFGVSLQFV SKVAVMDKAKPEFAKRVCSFMISFVIAGAGVTGLQMLKENQIEVEGRTEYKKA RNQRLPKVINEQWIKLOEKLKK"			
	repeat_region	8809..8842			
	repeat_region	/rpt_family="(GAAA)n"			
		9732..9767			
	repeat_region	/rpt_family="(TAAA)n"			
		complement(10971..11087)			
	repeat_region	/rpt_family="(CAAAA)n"			
		complement(11038..11085)			
	repeat_region	/rpt_family="POLY-A"			
	mRNA	Join(<11437..11628,11808..11867,12158..>12280)			
		/gene="At2g35090"			
	gene	<11437..>12280			
		/gene="At2g35090"			
		/note="T4C15.24; predicted by genscan"			
		Join(11437..11628,11808..11867,12158..12280)			
	CDS	/gene="At2g35090"			
		/note="hypothetical protein"			
		/codon_start=1			
		/protein_id="AAC61826.1"			
		/db_xref="GI:3668094"			
		/translation="MAMKRTISIVIRIGLSYRSLSSGNHVDKLGVMKALEAINVPS TQAEALTGATSGLESVVEKEIQSTVAEVRKMRADINKFNELKAEFTTKYEAK			

their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,

<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by repeatmasker (Arian Smit,

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6p23, F5j6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES

source

1. 122871

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="11"

/note="Sequence from clone T14P1"

/complement(519..627)

/rpt_family="TAAAAA)n"

join(2227..2322,2480..2527,2650..2705,3098..3150,

3250..3641)

/gene="At2g45170"

2227..3641

/gene="At2g45170"

/note="T14P1.2"

join(2488..2527,2650..2705,3098..3150,3250..3469)

/gene="At2g45170"

/codon_start=1

/product="putative microtubule-associated protein"

/protein_id="AAF18602.1"

/db_xref="GI:6598366"

/translation="MNKGSIFKMDDEKRAEAGRIREKVPDRTPVIVEAKSEVP

NIDKKYLPDGLFGQVYVIRKRIKLSAEKAFIFVDNVLPTIGELMSSVYEDKD

EDGFLYITISGENTFGASSI

/complement(2953..>122871)

/note="Sequence from clone F4L23"

4282..4870

/gene="At2g45180"

4282..4870

/gene="At2g45180"

/note="F4L23.31"

4320..4724

/gene="At2g45180"

/codon_start=1

/product="putative proline-rich protein"

/protein_id="AAB82643.1"

/db_xref="GI:2583134"

/translation="WASKALAVTALLITLNLFFTVSTKCPPTTPKPKTPKSPKK

APAVKPTPTDTKLGVGADLLGLVNVVVGSPKTPCCITLQGLANLAAVCLTALK

ANVLGINLVPTDITLLNLYCKGVPHGFQCS"

5379..5403

/rpt_family="POLY_A"

/complement(join(<6393..6470,6586..6663,7082..7157,

7248..7296,7872..8004,8099..8260,8382..>8495))

/gene="At2g45190"

/complement(<6393..>8495)

/gene="At2g45190"

CDS

/note="F4L23.30"

/complement(join(6393..6470,6586..6663,7082..7157,

7248..7296,7872..8004,8099..8260,8382..8495))

/gene="At2g45190"

/note="unknown protein"

/codon_start=1

/protein_id="AAB82644.1"

/db_xref="GI:2583135"

/translation="MSMSMSSPSSAVGSPDHFSPSDHLYVQCNFOTILAVNVYPT

SLFTVTYVRCGCCTNLLSVNMRSVLPASNLQQLQGLGPHSYFNPDILLEEDAPSNM

MMNMNOHTMNDIPSPMDLHQHEIPKAPVNRPRPEKROVPAYNRFIKKEIQRIK

AGNDISHREAFSAAKNAWHPHIFGLVPDNPVKTKNMPQEGEDNMVMEGEFYA

PAAAVGVITP"

10147..10201

/rpt_family="(CA)n"

12613..12646

/rpt_family="(CATA)n"

/complement(join(15613..15729,15821..15950,16066..16137,

16232..16341,16810..17006,17451..17647))

/gene="At2g45200"

/complement(15613..17647)

/gene="At2g45200"

/note="P4L23.29"

/complement(join(15632..15729,15821..15950,16066..16137,

16232..16341,16810..17006,17451..17583))

/gene="At2g45200"

/codon_start=1

/product="putative cis-Golgi SNARE protein"

/protein_id="AAB82642.1"

/db_xref="GI:2583133"

/translation="MTESLDLOESGWEELREARKIEGDLVDKSSYAKLGARFTQG

GYVDTGSPVSGRSWKSWEIQLLEKLLDINDSMSCRAAPTSVTKLARHR

DILHEYTQEFFRIKGNINSLEHAEKLLSVDKDISEYKASGMSFGVQVIREASIHG

SISHHDDVIGQAQTRAVIGSQSLFSDVQGVKNLGDKFPVIRGLLGSIRKRSRDT

LILSAVIAACTIFLIYWLK"

/complement(17986..18862)

/note="molecular marker ML"

/complement(18449..18564)

/rpt_family="(TAAAA)n"

19741..>20315

/gene="At2g45210"

19741..>20315

/gene="At2g45210"

/note="F4L23.28; member of auxin-induced protein multigene

family"

19827..20315

/gene="At2g45210"

/codon_start=1

/product="putative auxin-induced protein"

/protein_id="AAB82641.1"

/db_xref="GI:2583132"

/translation="MRKIIGFRIGRVSRIEKRRIORSYGNRIHSTOQACMLMRPL

AKLKGWGLKOSFRRTSRISAYIPVDHKKADPVRGHAIYVQKDGDCRHLVPI

VYFNHPLFGLREAEKEYFCHEGGITIPCLYSDFERVKTRIASGSSRVFPWRHC

RN"

/complement(join(<22224..22909,23488..>24334))

/gene="At2g45220"

/complement(<22224..>24334)

/gene="At2g45220"

/note="F4L23.27"

/complement(join(22224..22909,23488..24334))

/gene="At2g45220"

/codon_start=1

/product="putative pectinesterase"

/protein_id="AAB82640.1"

/db_xref="GI:2583131"

/translation="MAFRAYIINFVILCILTAVSTVSGYNQKDVKAWCSOTPNPKPCEY

FTLHNSNEPIKSESEFLKISMKLVLDRAITLAKTHAFTLGPCKRDTREAAEDCIKL

YDLTVSKINETMDPNVKCSKLDQTLSTALTNLDTCRAGFLGTVDITVLPLMSNNV

SNLICTLAINKVPNTTPPEKDPSPWPKGDKLQSTPKDNAAVAKDGSNGFT

IKKADAAAGSGRFVLYKQGVSENLEIRKKNVLMRGD IGT IITGSKVSGGGTTT

FNSATVAAGDGIAGITFRNTAGASQAVNLRSGSLSVFYQCSFAYODTLVH

SNRQFYRDCDVTGYDFIFGNAAVLQNCNIFARRPSKNTINTITAQGRSDPNQNTGII

```
IHNSRVTAASDLRPVLGKTYTLGRPMWRQSRVFMKTSLSLSDSLDPRGWLWDGNFAL
KLFYAEFQNTGPGASTSGRVTWPGFRVLGASASEASKFTVGTFLAGGSWIPSSVPFES
GL
repeat_region 23009..23038
/rpt_family="(TA)n"
repeat_region 28701..28741
/rpt_family="POLY_A"
repeat_region 29228..29228
/complement(29132..29228)
/rpt_family="POLY_A"

Query Match 21.1%; Score 205.2; DB 8; Length 122871;
Best Local Similarity 61.1%; Pred. No. 9.5e-47;
Matches 352; Conservative 0; Mismatches 218; Indels 6; Gaps 1;

Qy 58 taacatggaacttaacagatctgaagcagcagcaagcaagccgagacacactccacccg 117
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 105952 TGATATGATCCACAGACACTCTTACGCCGGAAGATCAAGTACTCTCTGCTCCG 106011

Qy 118 tggagccaccagctcagccagcagctctgctcttctccgagcgtcgtcagcgtgctg 177
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106012 TGGAGAAAGCCGCGCGGAGGAGGAGAGATAATCATCATCAGAGAGCCACGTGGCAG 106071

Qy 178 tctgcaggttcccaaaacacacacccacccacccgacattataactagagatagctctaa 237
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106072 ACCACGCGGATCTAAGAACAAACCAACCCGCAATCATCATCATCTGAGACCGCAAA 106131

Qy 238 cgtccttgatcagcagctcttgaagtcacccctccggttcgagacatataccgagcgagcttc 297
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106132 CGCTCTCAAAATCTCATGTGTCATGGAAGTAGCAACAGGATGTGACGTCATGGAAGTGTAC 106191

Qy 298 caccatcagccactcgtcgcgcgtcgcgtctgcatataagcgcagcagcgtgagctcac 357
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106192 CGTCTCTGCTCGCGTCCGCAACGTGGCATCTGCTTTGAGCGGAAACGCGCGGTAC 106251

Qy 358 taacgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 411
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106252 CAACGTTACCATAGACAAACCAACGAGTTCAGTACCTGGTGGTGTCTGCTGTTAACTT 106311

Qy 412 gcagtcgctgttgaacatttctgtttgacgcgtgacgcgtgacgcgtgacgcagcagc 471
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106312 ACAGGAGCTTTCAGAGATCTTCTCTCTCGGATCATCTCTCTCTCGGCTCCACC 106371

Qy 472 gggagcagaggtttgacggtgtatctagccgaggttcaagacaaagtgttagagagga 531
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106372 AGCTGCGTCAGCTTAAGGATTTACTAGCGGTGTCAGGACAGGTTGTTGAGGAGAG 106431

Qy 532 tgtggtggttcgttaattgtcttcggagcagcagcagcagcagcagcagcagcagcagc 591
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106432 CGTGTGTCCTCACTCATGCTTTCAGGACCTGTAGTGATTATGTCAGCTTCGTTTGAAA 106491

Qy 592 cgcagtttatgataggttaccgattgaagaggaaga 627
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106492 CGCTGCGTATGAGAGACTGCGGTGGAGGAAGACGA 106527

RESULT 7
AC015450
LOCUS
DEFINITION Arabidopsis thaliana chromosome 1 BAC F1466 genomic sequence,
complete sequence.
ACCESSION AC015450
VERSION AC015450.5 GI:12323968
KEYWORDS HTG.
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 100806)
Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC F1466 genomic sequence
```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 100806)
Lin,X. and Kaul,S.
Direct Submission
Submitted (16-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 100806)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280761.
Address all correspondence to:at@tigr.org

BAC clone F1466 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GenemarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/Genemark/), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
http://www.tigr.org/soflab/glimmerm.htm/glimmerm.html, and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES
Source

1. 100806
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F1466"

misc_feature

1
/note="35000 nt before this point were not included in
the submitted sequence, due to overlap with another BAC
[F15M4]"
-5217..>5774
/gene="F1466.1"
5217..5774
/gene="F1466.1"
/note="contains Pfam profile: PF00097 Zinc finger, C3HC4
type (RING finger)"
5217..5774
/gene="F1466.1"
/codon_start=1
/product="putative RING zinc finger protein; 5217-5774"
/protein_id="AAG51946.1"
/db_xref="GI:12323975"
/translation="MARLLFRLLEANSSTPAEASPPFNSDLVILAVLICALTCIIG
LIAVSRCAWLRRIASRNQSDQTHPPVAANKGLKKVLRSLKLTYSPPSPPAEKL
ECAICLTFEAGDELRLVQSGHGFHVSICDITWLGSHSSCPCSQIILVVTCHRCGGL
PGSSSSGPEPDTRIKQREDGPDNL"
complement(join(<8331..8855,9526..9800,9933..>10137))
/gene="F1466.2"
complement(8331..10137)
/gene="F1466.2"

CDS

5217..5774
/gene="F1466.1"
/codon_start=1
/product="putative RING zinc finger protein; 5217-5774"
/protein_id="AAG51946.1"
/db_xref="GI:12323975"

mRNA

5217..5774
/gene="F1466.1"
/note="N-term similar to N-term of NAM GB:CAA63101
[Petunia x hybrida] (apical meristem formation), CUC2
GB:BAAL19529 [Arabidopsis thaliana], GRAB2 protein
GB:CAA03372 [Triticum sp.]
complement(join(8331..8855,9526..9800,9933..10137))
/gene="F1466.2"
complement(8331..10137)
/gene="F1466.2"

gene

5217..5774
/gene="F1466.1"
/note="N-term similar to N-term of NAM GB:CAA63101
[Petunia x hybrida] (apical meristem formation), CUC2
GB:BAAL19529 [Arabidopsis thaliana], GRAB2 protein
GB:CAA03372 [Triticum sp.]
complement(join(8331..8855,9526..9800,9933..10137))
/gene="F1466.2"
complement(8331..10137)
/gene="F1466.2"

CDS

5217..5774
/gene="F1466.1"
/codon_start=1
/product="putative RING zinc finger protein; 5217-5774"
/protein_id="AAG51946.1"
/db_xref="GI:12323975"

Db	54066	AAACAAGACAGCGCAAAACCGCTTCGAACTCAGCTCATGGAGATAGGACGCGATGTGA	54007
QY	280	catatccgagggcaggtctccaccctacgcacccatctcgccggtcgccggtttggtcattcaag	339
Db	54066	CATAGTTGACTGTATGGCTTACGTTGCGTAGACGCCGCCAAAGAGGCGTTTTCGTTATGAG	53947
QY	340	cgcacgaggtcggtctcaactaacgctcacgatacagcaacctcgccgctccggtggtgaggg	399
Db	53946	CGGTACAGGAAGCGTTACTTACGTCACTATACGTACGCCCTGGATGCCACC---TGGCTC	53890
QY	400	tgtgattacccctgcattggtcggtttgacattttgtcttttgcacggtagctcgctccacc	459
Db	53889	GGTGTGTAGCTTACGCGCGGTTTGAATCCTCTCTCTTTCCGGGATCTTTCTTTCGCTCC	53830
QY	460	gctgcaccaccgggagcagaggtttgacggtgtatctacgacgaggtcaagcacagt	519
Db	53829	GCTCGCCCGCGCTGCACGCCACCGGACTTAAGCGTTTACTACCGGAGGACAAGGCGAGGT	53770
QY	520	tgtagaggaagtgtggttgcgttaattgcttcgggaccggttagttgttatggtgtgc	579
Db	53769	CGTTGAGAGTGTGTGCTGGGACCTTTGTTGTGTTTCGSGTCTCTGTGGTGTATATGGCGGC	53710
QY	580	ttctttgcaaacgcagtttatatgattaccgattgaagaggagaagaaacccaccgcc	639
Db	53709	TTCTTTTGAATACGCGCGTACGAAGGCTGCCTTTTGGAAAGATGAGATGCAGACGCC	53650
RESULT	9		
LOCUS	ATCHRIV33/c		linear PLN 16-MAR-2000
DEFINITION	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 33.		
ACCESSION	AL161533		
VERSION	AL161533.2	GI:7267889	
KEYWORDS			
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 2832 to 3389); 3666 to 4213)		
AUTHORS	Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 123056 to 190026)		
AUTHORS	Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 34336; 23831 to 131430)		
AUTHORS	Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
JOURNAL	Unpublished		
REFERENCE	4 (bases 1 to 190026)		
AUTHORS	EU Arabidopsis sequencing project.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-8152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk		
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATCHRIV32 at the 5' end and an overlap with ATCHRIV34 at the 3' end.		

```
FEATURES      Location/Qualifiers
source        1..190026
              /organism="Arabidopsis thaliana"
              /varietal="Columbia"
              /db_xref="taxon:3702"
              /chromosome="4"
              5417..6920
gene
```

```

/feature="AT4g11890"
5417..5486
/feature="AT4g11890"
/number=1
join(5417..5486,5584..5675,5763..5973,6065..6302,
6396..6534,6612..6920)
/feature="AT4g11890"
/note="strong similarity to KI domain interacting kinase 1
-zea mays, PR2:702053
Contains Protein kinases signatures and profile
AA44-66;Protein kinases signatures and profile AA159-171"
/codon_start=1
/product="protein kinase-like protein"
/protein_id="CAB78232.1"
/db_xref="GI:7267890"
/translation="MAVTSLLDTVFRRRKKKSTEFISSVFEDLTIKAATNDFSELV
GRGFGVYVGRQNGEIAVKILSTSSIRTEQFHNELIILSKLKHNLINLLGECT
KRDQGLVYVEFMPNSLDLCFLDPHRAQLNWMCRNIIDGIARGLRYLHEESGLNVV
HRDKKPNLLDSELDPKYVGFELAPMQQGENAETETVIGVGLDPEYIRSGRVS
VKSDVYAFGVLTILTIISRRKAWSDGDSLIKYVRRCWNRGEADIVHEVNREREYS
ISELRVITHALLCVDNAERPRNIDKVLHWFSCFSTPLDPDTFGNREFVEETNPW
SPSLSPGHSSVTSPISSR"
5487..5583
/feature="AT4g11890"
/number=1
5584..5675
/feature="AT4g11890"
/number=2
5676..5762
/feature="AT4g11890"
/number=2
5763..5973
/feature="AT4g11890"
/number=3
5974..6064
/feature="AT4g11890"
/number=3
6065..6302
/feature="AT4g11890"
/number=4
6303..6395
/feature="AT4g11890"
/number=4
6396..6534
/feature="AT4g11890"
/number=5
6535..6611
/feature="AT4g11890"
/number=5
6612..6920
/feature="AT4g11890"
/number=6
7389..10690
/feature="AT4g11900"
complement(7389..7679)
/feature="AT4g11900"
/number=1
complement(join(7389..7679,7778..7928,8208..8445,
8535..8765,8889..9043,9121..9255,9322..10690))
/feature="AT4g11900"
complement(join(7389..7679,7778..7928,8208..8445,
8535..8765,8889..9043,9121..9255,9322..10690))
/note="strong similarity to KI domain interacting kinase 1
-zea mays, PID:g2735017
Contains Prokaryotic membrane lipoprotein lipid attachment
site AA472-482;Protein kinases signatures and profile
AA658-670
contains EST gb:AI992820.1, T20644, AA586217"
/codon_start=1
/product="KI domain interacting kinase 1-like protein"
/protein_id="CAB78233.1"
/db_xref="GI:7267891"

```

```

/translation="MQICKKNVFLLYGVLFVLSFQVSSSTDTISTNQPLSGFETIVS
DGNIFELGLFTPTDITYDHRNYYIGMYRHSVPQIVVMANRESPLGDASTYLLKIL
DGNLILHDNISATRSKTTEGTSRRSPQKISEGNLLFHEFTVMSGTVNSMSKDVQAVLF
EDNKLHSLVTVMNRKSKVMSGPLYDMLQSPKGPPELQGTKLSPTLNDSEYITFSV
DQOSRYLVMSVSGQFMQLQVHVVDLQSHRYVILSQPDNRCDVYNSCGSGICNENREPP
PCRVPGFRKREFSQGSDSDSNQYSGGCKRETYLHCYKRNDDELPIENMKLATDPTTASV
LFSGTFRTCAKSCVADCSQAYANDGNKCLVMTKDAFNQLQDLANKHGTFFLRASSN
ISTANNRKTESKGSIVLPLVLASIVATAACFVGLYCCISRIIRKKQRDEKHSRE
LLEGLIDIDAGNMCVNLHDIMVATNSFRKKKLGEGGPGVYKGLPNGMEVAIKR
LSKSSOGITFEKNEVLLIKLQHKNLVRLGVCYEGDEKLLIYVMSNKSJLGLLFD
SLKSELDEWETRKIVNGTTRGLOVLYHEYSRLRIHRBLKASNILLDEMMPKSDFG
TARIFECQIDDSQIRIVTFGYMSPEYALGVSEKSDIYSFGVLLLEIISGKKATR
FVNDQKHSIAIYEWESMCEKGVSIIDPEMCCSYSLSEAMRCIHALLCYQDHPKDR
PMISQIVYVMSLNDNLTLPKQPTFSNVNLNGOQLDYVFSINEATOTELEAR"
complement(7680..7777)
/feature="AT4g11900"
/number=1
complement(7778..7928)
/feature="AT4g11900"
/number=2
complement(7929..8207)
/feature="AT4g11900"
/number=2
complement(8208..8445)
/feature="AT4g11900"
/number=3
complement(8446..8554)
/feature="AT4g11900"
/number=3
complement(8555..8765)
/feature="AT4g11900"
/number=4
complement(8766..8888)
/feature="AT4g11900"
/number=4
complement(8889..9043)
/feature="AT4g11900"
/number=5
complement(9044..9120)
/feature="AT4g11900"
/number=5
complement(9121..9255)
/feature="AT4g11900"
/number=6
complement(9256..9321)
/feature="AT4g11900"
/number=6
complement(9322..10690)
/feature="AT4g11900"
/number=7
13583..16588
/feature="AT4g11910"
13583..13711
/feature="AT4g11910"
/number=1
join(13583..13711,13978..14151,14239..14406,14643..14879,
15546..15719,15839..15996,16238..16588)
/note="similarity to hypothetical protein F7H19.100
-Arabidopsis thaliana, PID:el310060
Contains Zinc finger, C2H2 type, domain AA420-440
contains EST gb:234017, 234018"
/codon_start=1
/product="putative protein"
/protein_id="CAB78234.1"
/db_xref="GI:7267892"
/translation="MCSLATNLLLPKMKPVPEKLTSSLCVTRRSKMKNRSTVPV
ARLGFALFEASKLVFLGVDEKHKAPKLPRTYTLTHSDITAKLTALISINNSQL
QGMANKLRDEVYEMKVKGMKSLHVCCHISGGHFFLNLAJLAKRYLIFCKELPVVLE
APAHGDEVLNNHPELQSPVWVYIPYKNVCEKGLPEAKMSQHQHGRTHKKS
FTLPBPCDECKCCFPVAVRLFQGAIFEASKLVNFKLVGVDEKHPNLPRTYTFTHS
DITAKLTALSHSINNSQLQGWANRLYRDEVVAEWRKVKSNKSLHVCCHISGDHFLD

```


DEFINITION	Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
ACCESSION	No. 1.
VERSION	Z97336
KEYWORDS	Z97336.1 GI:2244788
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS	1 (bases 1 to 206606) Bevan.M., Stiekema.W., Murphy.G., Wambutt.R., Pohl.T., Terry.N., Kreis.M., Kavanagh.T., Entian.K.D., Rieger.M., James.R., Puigdomenech.P., Hatzopoulos.P., Obermaier.B., Duesterhoft.A., Jones.J., Palme.K., Ansorge.W., Delseny.M., Bancroft.I., Mewes.H.W., Schueller.C. and Chalhatzis.N.
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 206606)
AUTHORS	EU Arabidopsis sequencing project.
TITLE	Direct Submission
JOURNAL	Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG. E-mail: schuellemips.biochem.mpg.de,mayermips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATPCA0 at the 5' end and an overlap with ATPCA2 at the 3' end. Location/Qualifiers 1..206606 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 1..26121 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone="BAC TAMU8H22" complement(join(685..744,829..903,1053..1232,1804..1806)) /gene="dl3200c" 685..1806 /gene="dl3200c" complement(685..744) /gene="dl3200c" /number=1 complement(join(685..744,829..903,1053..1232,1804..1806)) /gene="dl3200c" /note="strong similarity to ribosomal protein L41, Candida maltosa" /codon_start=1 /product="ribosomal protein" /protein_id="CAB10211.1" /db_xref="GI:2244789" /db_xref="SWISS-PROT:O23290" /translation="MVNIPKTKNTYCKNKECKKHTLHKVTQYKKGKSLAAQGRKRYD RKQSGYGGQTKPVFHKAKTKKIVLRQCQCKHFSQRPKRCRKHFEIGDKKGGKT SLF" complement(745..828) /gene="dl3200c" /number=1 complement(829..903) /gene="dl3200c" /number=2 complement(904..1052) /gene="dl3200c" /number=2 complement(1053..1232) /gene="dl3200c" /number=3 complement(1233..1803)
FEATURES	source
source	1..206606
gene	complement(join(685..744,829..903,1053..1232,1804..1806)) /gene="dl3200c"
gene	685..1806 /gene="dl3200c"
exon	complement(685..744) /gene="dl3200c"
CDS	complement(join(685..744,829..903,1053..1232,1804..1806)) /gene="dl3200c" /note="strong similarity to ribosomal protein L41, Candida maltosa" /codon_start=1 /product="ribosomal protein" /protein_id="CAB10211.1" /db_xref="GI:2244789" /db_xref="SWISS-PROT:O23290" /translation="MVNIPKTKNTYCKNKECKKHTLHKVTQYKKGKSLAAQGRKRYD RKQSGYGGQTKPVFHKAKTKKIVLRQCQCKHFSQRPKRCRKHFEIGDKKGGKT SLF"
intron	complement(745..828) /gene="dl3200c"
exon	complement(829..903) /gene="dl3200c"
intron	complement(904..1052) /gene="dl3200c"
exon	complement(1053..1232) /gene="dl3200c"
intron	complement(1233..1803)
exon	complement(1804..1806) /gene="dl3200c"
exon	complement(1807..1808) /gene="dl3200c"
exon	2229..2837 /gene="dl3205w"
gene	2229..7384 /gene="dl3205w"
CDS	join(2229..2837,2914..3069,3161..3268,3358..3468,3569..4957,5000..5212,6340..6428,6469..6560,6706..6749,7316..7384) /gene="dl3205w"
	/note="similarity to kinesin osm-3, Caenorhabditis elegans contains EST gb:496806" /codon_start=1 /product="kinesin like protein" /protein_id="CAB10212.1" /db_xref="GI:2244790" /db_xref="SPTREMBL:O23291" /translation="WAPTPSSSRNTOYTLIRTQTKORLNFHSHKTPNPGSKDPSP PEPVEVIGRIDIPDRKEKSPSILQVNTDQTVYRADYGVYRDTLDGVSESEGL BEFYKFFIERRIKGVKGNKCTIMMIGTGAAGSHTEGCGKEPQIVIRSLRDLGDS DDQGTFFVQTVLEVINEEYDLILSTSSNLLGIGPKGASTKRVLYGVMKKAKNASF ISGTEAGKISKEIVKERRIVKSTLCNERSRSHCIILIDVTVGGRLMLVQDSFEDD NIDAGQTGFKAQWOTAKINQNIATKRVSIIANGSDHVPFRDSKLTMLLQDSFEDD KSKTILMIICASDPKEMHKTCTLEYGAKAKICVRGSHTPNKRKYGGDESASAVILGS RIAMDERFIILQSEKKQKKEKNEAKOLKKEEVAALSLLTQREACATNEEEIK BEVNRSTQLLSKELDKLECEKRAEEFEVEMRRMEERIVQQQEELEMMRRLEIEI VEFRRNSGVSDETSFGAKRLRSYSDDDPGVMKMSDMDGDPKPVKQVWGAVSQSS NTISSFTNLQPKPSNMLTQMPYDRCLSTVFEEEEEVEEKEKIVVEDKSLCILT PMPSLNSEGLKENCENGADDEKASAKENNISVDQENGQDIIYVKEWTAADNPKLIT ASPDKNOFFSITNKALAVEAKENNISVDQENGQDIIYVKEWTAADNPKLIT TLRYTKDATLADLRKLEIYLGSDNQAFTELKLVINLNOQAOKAFHYFLPPECAOV FPLNSRADADEPGLVLDITPLFEVVLGFTLVLELAVVNDRIAG"
intron	2838..2913 /gene="dl3205w"
exon	2914..3069 /gene="dl3205w"
intron	3070..3160 /gene="dl3205w"
exon	3161..3268 /gene="dl3205w"
intron	3269..3357 /gene="dl3205w"
exon	3358..3468 /gene="dl3205w"
intron	3469..3568 /gene="dl3205w"
exon	3569..4957 /gene="dl3205w"
intron	4958..4999 /gene="dl3205w"
exon	5000..5212 /gene="dl3205w"
intron	5213..6339 /gene="dl3205w"
exon	6340..6428

RESULT 13	ATFCA9	81835 bp	DNA	linear	PLN 28-JUN-1999
ATFCA9/c	LOCUS				
DEFINITION	DEFINITION				
	Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment				
	No. 9.				
ACCESSION	Z597344				
VERSION	Z597344.2	GI:5281025			
KEYWORDS					
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta.				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	rosids; order II. Brassicales; Brassicaceae; Arabidopsi.				


```

Db 61170 GTGGTTGAGGAACGGTTGTGGGTAGCTTACGGCGGTGACCGGTCATTTGTGATTGCA 61111
      || || ||||| || || ||||| || || ||||| || || ||||| || || ||||| || || ||||| ||
Qy 578 gctcttttgcacacgagtttatgataggttaaccgattgaagaggaaga 627
      ||||| ||||| || || ||||| || || ||||| || || ||||| || || ||||| || || ||||| ||
Db 61110 GCTTCGTTTACTAATGTTGCTTATGAGAGACTTCTCTTTAGAAGAAGATGA 61061
      || || ||||| || || ||||| || || ||||| || || ||||| || || ||||| || || ||||| ||

RESULT 14
ATCHRIV47/c 198067 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 47.
ACCESSION AL161547
VERSION AL161547.2 GI:7268521
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 77203 to 176024)
AUTHORS Rieger,M., Mueller-Auer,S., Zipp,M., Schaefer,M., Mewes,H.W.,
Lemcke,K. and Mayer,K.F.X.
JOURNAL Unpublished
REFERENCE 2 (bases 173478 to 198067)
AUTHORS Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
JOURNAL Unpublished
REFERENCE 3 (bases 76382 to 77642)
AUTHORS Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
JOURNAL Unpublished
REFERENCE 4 (bases 1 to 198067)
AUTHORS EU Arabidopsis sequencing,project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV46 at the 5' end and an
overlap with ATCHRIV48 at the 3' end.
FEATURES
Location/Qualifiers
1..198067
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
6476..7456
/gene="AT4g17690"
6476..7456
/gene="AT4g17690"
/note="strong similarity to peroxidase (EC 1.11.1.7),
Medicago sativa
Contains Peroxidases signatures AA186-196;Peroxidases
signatures AA57-68
contains EST db:AT990391.1, AI994341.1, AI994675.1,
AI995254.1, AA721954, R90036, 235219, AA713329"
/codon_start=1
/product="peroxidase like protein"
/protein_id="CAB78772.1"
/db_xref="GI:7268522"
/translation="MSSIVNVLVFLVFLVPSIYAPPNLTKDYQKCPDFNKIVRE
TVPKGGQPTTAAGTLRLFLHDFHDCFMGCDASVLATNSFNKAERDDDLNESLPGDAF
DIVTRIKTALELSCGVVSCADILAQATRLDVTWVGPPFVEVKGKRGDGFESKAHKV
GNLPLANGSVFMDLSIFKNGFTLKLVALSGGHTIGFSHCKEFSNRPFPKVDPELNA
KFAVGLKDLCKNFETNKTMAAPLDVPTPKFDNMYFKNLKRGGLIASDHTLFKDPST
RPFVELYANNOTARTAFEDFARAMEKLGVRGVCKEKGDEVRRCRDHFENKLVN"
6476..7456
      * exon

```

```

/gene="AT4g17690"
/number=1
repeat_region
repeat_region
exon
gene
CDS
join(14671..14712,14892..15086,15130..15360,15665..15754,
15846..15952,16639..17032)
/note="similarity to T08A11.2, Caenorhabditis elegans
Contains prokaryotic membrane lipoprotein lipid attachment
site AA67-77"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB78773.1"
/db_xref="GI:7268523"
/translation="MEETFLVRVREAGGFIEHMDTKPSSCVIGCFEFDNDYNA
HIVLFARMGTHKYNMIOVKNLSPLVYICSSSSCKRGTNLQSCVEKYNRSPRYSAF
YITLIADPDACNSLVTFTQTRVVEGFNMTKLSNIAKPKPGQKGLGTLTYQRYKKS
ELTHDWIRLYMELAFKAHLRSPKNLSKLVIVMVAIETKDNMEPSEKRNAAF
FYINCDYLPKYPNLLITDVTVDERSGSDTVDGANTVIDDGGKIESFGDVPQDL
XCNSEQARRDTGTALDTIGWIHNNRESNGKWYFRRIFAVHGLNSDGVGRVHQE
TPGVIGACFLKHQLYIMR"
14713..14891
/gene="AT4g17700"
/number=1
14892..15086
/gene="AT4g17700"
/number=2
15087..15129
/gene="AT4g17700"
/number=2
15130..15360
/gene="AT4g17700"
/number=3
15361..15664
/gene="AT4g17700"
/number=3
15665..15754
/gene="AT4g17700"
/number=4
15755..15845
/gene="AT4g17700"
/number=4
15846..15952
/gene="AT4g17700"
/number=5
15953..16638
/gene="AT4g17700"
/number=5
16639..17032
/gene="AT4g17700"
/number=6
complement(join(16676..17044,17299..17534,17599..17935,
18095..18193,18280..18639,18716..18868,19016..19133,
19241..19404,19488..19637))
/gene="AT4g17710"
complement(join(16676..17044,17299..17534,17599..17935,
18095..18193,18280..18639,18716..18868,19016..19133,
19241..19404,19488..19637))
/gene="AT4g17710"
/note="similarity to homeotic protein GL2, Arabidopsis
thaliana
Contains 'Homeobox' domain signature and profile
AA121-144"
/codon_start=1
/product="GLABRA2 like protein"
/protein_id="CAB78774.1"
/db_xref="GI:7268524"
/translation="METKDKKEKGHMLVNSDNVFGSVSSPTTTTIONPNFTSFENPN"

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 20:07:47 ; Search time 231.22 Seconds
(without alignments)
7232.396 Million cell updates/sec

Title: US-09-823-676-1
Perfect score: 974
Sequence: 1 cccccgacctgctctaca.....cgaaatgcgaattagggtt 974

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	974	100.0	974	22	Arabisopsis thalia
2	974	100.0	974	22	Nucleotide sequenc
3	196.8	20.2	1020	22	A. thaliana transc
4	191.6	19.7	1307	21	Arabisopsis thalia
5	191.6	19.7	1310	21	Arabisopsis thalia
6	190.4	19.5	1097	21	Arabisopsis thalia
7	179.2	18.4	1192	21	Arabisopsis thalia
8	178.8	18.4	1190	21	Arabisopsis thalia
9	178.8	18.4	1251	24	Arabisopsis cdna e

c	10	177.2	18.2	1303	21	AAC35116	Arabisopsis thalia
	11	127.6	13.1	12319	21	AA289379	A. thaliana NCW1 g
	12	42	4.3	1219	21	AAC44421	Arabisopsis thalia
	13	42	4.3	1263	22	AAC05779	Arabisopsis thalia
	14	42	4.3	1387	21	AAC38928	Arabisopsis thalia
	15	42	4.3	1606	22	AAD06652	A. thaliana transc
	16	38.6	4.0	1745	22	AAI29497	C883P determined c
	17	37.4	3.8	1361	22	AAD02889	Dinorphototheca sinu
	18	37	3.8	7696	22	AAS46399	Tumour suppressor
	19	36	3.7	3062	23	ABL23912	Drosophila melanog
	20	35.8	3.7	510	23	ABL22333	Drosophila melanog
	21	35.8	3.7	1407	22	AAD06671	A. thaliana transc
	22	35.8	3.7	2606	23	ABL22332	Drosophila melanog
	23	35.8	3.7	4280	18	AAV74392	Staphylococcus aur
	24	35.6	3.7	576	22	AAS01677	Human apolipoprote
	25	35.2	3.6	5193	23	ABL23107	Drosophila melanog
	26	35.2	3.6	9577	23	ABL23106	Drosophila melanog
	27	34.8	3.6	330	24	AAS97212	Neisseria meningit
	28	34.8	3.6	330	24	AAS97230	DNA encoding a gly
	29	34.8	3.6	1039	18	AA783800	Neisseria meningit
	30	34.8	3.6	5276	22	AAD08671	Drosophila sp. mus
	31	34.6	3.6	7320	23	ABL24159	Drosophila melanog
	32	34.6	3.6	9718	23	ABL24158	Drosophila melanog
	33	34.6	3.6	16831	22	AAK73455	Human immune/haema
	34	34.4	3.5	4409	23	ABL17895	Drosophila melanog
	35	34.4	3.5	5021	23	ABL11783	Drosophila melanog
	36	34.4	3.5	7394	23	ABL11784	Drosophila melanog
	37	34.4	3.5	7614	23	ABL11782	Drosophila melanog
	38	34.2	3.5	781	23	ABL07225	Drosophila melanog
	39	34.2	3.5	1075	22	ABA48937	Human breast cell
	40	34.2	3.5	1075	22	ABA66852	Human foetal liver
	41	34.2	3.5	1075	22	ABA33921	Probe #12387 for g
	42	34.2	3.5	1075	22	AAK15288	Human brain expres
	43	34.2	3.5	1075	22	AAK41011	Human bone marrow
	44	34.2	3.5	1075	22	AAI21781	Probe #11714 for g
	45	34.2	3.5	1075	22	AAI47064	Probe #15750 used

ALIGNMENTS

RESULT	1
AA05839	
ID	AA05839 standard; cdna; 974 BP.
XX	
AC	AA05839;
XX	
DT	31-JUL-2001 (first entry)
XX	
DE	Arabisopsis thaliana transcription factor G1073 cdna.
XX	
KW	Transcription factor; trait modification; structural characteristic;
KW	developmental characteristic; gene therapy; agricultural biotechnology;
XX	
OS	Arabisopsis thaliana.
XX	
FH	Key
FT	Location/Qualifiers
FT	62..874
FT	/*tag= a
FT	/product= "Transcription factor G1073"
XX	
PN	WO200136444-A1.
XX	
PD	25-MAY-2001.
XX	
PF	14-NOV-2000; 2000WO-US31325.
XX	
PR	17-NOV-1999; 99US-0166228.
PR	17-APR-2000; 2000US-0197899.
PR	22-AUG-2000; 2000US-0227439.
XX	
PA	(MEND-) MENDEL BIOTECHNOLOGY INC.

PA (RIEC/) RIECHMANN J L.
PA (REUB/) REUBER L.
PA (KEDD/) KEDDIE J.
PA (RATC/) RATCLIFFE O.
PA (HEAR/) HEARD J.
PA (SAMA/) SAMAHA R.
PA (YUGG/) YU G.
PA (JIANG/) JIANG C.
XX
XX Riechmann JL, Reuber L, Keddle J, Ratcliffe O, Heard J, Samaha R;
PI Yu G, Jiang C;
XX
XX WPI: 2001-335996/35.
DR P-PSDB; AAE01961.
XX
XX Nucleic acids encoding plant transcription factor polypeptides, useful
PT for altering the developmental and structural characteristics of
PT plants, e.g. corn, potato and cotton plants -
XX
XX Claim 4; Page 102-103; 140pp; English.
XX
XX The present sequence is Arabidopsis thaliana transcription factor G1073
CC cDNA. The transcription factors are used to modify traits associated
CC with structural or developmental characteristics of plants, e.g. soybean,
CC wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa,
CC sugar cane, turf, banana, blackberry, blueberry, strawberry, raspberries,
CC cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes,
CC honey dew, lettuce, mango, melon, papaya, peas, watermelon, pineapple,
CC spinach, squash, sweet corn, tobacco, tomato, peppers, onion, roseaceous
CC fruits and/or vegetable brassicas when their expression levels are
CC altered. The manipulation of transcription factor levels in plants
CC offers great potential in agricultural biotechnology for modifying
CC plant's traits. The transcription factors are also used in gene therapy.
XX
XX Sequence 974 BP; 222 A; 246 C; 279 G; 227 T; 0 other;

Query Match 100.0%; Score 974; DB 22; Length 974;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cccccgacctgctctacagagacctgaagattccagaccacccacctgatacaaaaataa 60
DB 1 cccccgacctgctctacagagacctgaagattccagaccacccacctgatacaaaaataa 60
QY 61 catggaacttaacagatctgaagcagacgaagcaaaagccgagaccactccaccgggtg 120
DB 61 catggaacttaacagatctgaagcagacgaagcaaaagccgagaccactccaccgggtg 120
QY 121 agccaccagctcagccagcctctggtctctctcccgagcgtcgtccacgtggtcgtcc 180
DB 121 agccaccagctcagccagcctctggtctctctcccgagcgtcgtccacgtggtcgtcc 180
QY 181 tgcaggttccaaaacaaacccaaactccgacgattataactagagatgctcctaact 240
DB 181 tgcaggttccaaaacaaacccaaactccgacgattataactagagatgctcctaact 240
QY 241 ccttagatcacacgtttctgaagtcacctccggttcgacatatacggagcagtcctccac 300
DB 241 ccttagatcacacgtttctgaagtcacctccggttcgacatatacggagcagtcctccac 300
QY 301 ctacgcactcgtcgcggctcgggtttgacattataaagcggcagcgggtgcggtcactaa 360
DB 301 ctacgcactcgtcgcggctcgggtttgacattataaagcggcagcgggtgcggtcactaa 360
QY 361 cgtcagatcacgacacccctcgcgtcgtggtgaggtgtgattaccctgcagtgctg 420
DB 361 cgtcagatcacgacacccctcgcgtcgtggtgaggtgtgattaccctgcagtgctg 420
QY 421 gtttgacattttgtctttgaccggtactcgcgttccaccgctgcacaccgagcagcagg 480
DB 421 gtttgacattttgtctttgaccggtactcgcgttccaccgctgcacaccgagcagcagg 480

QY 481 aggtttgacggtgttatctagcggaggtcaaggacaaagttaggaggaatgtgctgg 540
DB 481 aggtttgacggtgttatctagcggaggtcaaggacaaagttaggaggaatgtgctgg 540
QY 541 ttcgttaattgcttcgggaccggttagtgtgagtgctgttctttttgcaaacgcagttca 600
DB 541 ttcgttaattgcttcgggaccggttagtgtgagtgctgttctttttgcaaacgcagttca 600
QY 601 tgatagttaccgattgaagagaaagaaacccaccgagaaacccaccgaggtgagca 660
DB 601 tgatagttaccgattgaagagaaagaaacccaccgagaaacccaccgaggtgagca 660
QY 661 gcagcagcggagggcgtctcagtcgtaggttaccggaggtggggcccgagcgtgtga 720
DB 661 gcagcagcggagggcgtctcagtcgtaggttaccggaggtggggcccgagcgtgtga 720
QY 721 gtcaaacctccaagtggaatggtaggaggtgtgtcttctacaatcttggaatgaa 780
DB 721 gtcaaacctccaagtggaatggtaggaggtgtgtcttctacaatcttggaatgaa 780
QY 781 tatgaacaattttcaattctccggggagagatatattacggtatgagcgggtgagcagg 840
DB 781 tatgaacaattttcaattctccggggagagatatattacggtatgagcgggtgagcagg 840
QY 841 aggtggcggcgtgcagctagaccgcggttttagagtttttagcgtttttgtaacaccttt 900
DB 841 aggtggcggcgtgcagctagaccgcggttttagagtttttagcgtttttgtaacaccttt 900
QY 901 tgttgcgtttgcgtgtttgacctcaactactaggtactagctatagcgtttgcgaaat 960
DB 901 tgttgcgtttgcgtgtttgacctcaactactaggtactagctatagcgtttgcgaaat 960
QY 961 gogaatattaggtt 974
DB 961 gogaatattaggtt 974
RESULT 2
AAF80408
ID AAF80408 standard; cDNA; 974 BP.
XX
XX AAF80408;
DT 29-JUN-2001 (first entry)
XX
DE Nucleotide sequence of a plant transcription factor G1073.
XX
KW Transcription factor; flowering time; transgenic plant; vernalisation;
KW plant development; plant physiology; flowering; ss.
XX Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
CDS 62..874
FT /*tag= a
FT /product= "transcription factor"
XX
XX WO200126459-A2.
XX
XX 19-APR-2001.
XX
XX 11-OCT-2000; 2000WO-US28141.
XX
XX 12-OCT-1999; 99US-0159464.
PR 08-NOV-1999; 99US-0164132.
PR 17-NOV-1999; 99US-0166228.
PR 17-APR-2000; 2000US-0197899.
PR 22-AUG-2000; 2000US-0227439.
XX
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (RATC/) RATCLIFFE O.
PA (HEAR/) HEARD J.
PA (SAMA/) SAMAHA R.


```
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151130.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161004.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 19.7% Score 191.6; DB 21; Length 1310;
Best Local Similarity 64.7%; Pred. No. 2.4e-50;
Matches 303; Conservative 0; Mismatches 159; Indels 6; Gaps 1;

QY 161 cgtcgtccacgtggtcgtcctgcagggttccaaaacaaacccaaacccctccacacattata 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 cgtagaccacagaggtagaccaccaggatccaaaacaaacccaaacccctccacattttgtg 416

QY 221 actagagatagcttaacgtctcttagatcacacgttcttgaagtcacccctcggttcggac 280
    || |||| || || |||| || || || || || || || || || || || || || || || ||
DB 417 acaagagacagcccaacgcactccgttagccatgctctggagatctccgacggcagtgac 476

QY 281 ataccggaggagctctccacactacgcactctcctcaagacgagcagcggttgcgttcacgc 340
    || |||| || || || || || || || || || || || || || || || || || || || ||
DB 477 gtcgcgacacaaatcgctcactctcctcaagacgagcagcggttgcgttcacgc 536

QY 341 ggcacgggtgcgtcactaaacgtcacgatacggcaacactcggctcggctcgggtggaggt 400
    || || || || || || || || || || || || || || || || || || || || || || ||
DB 537 gggacaggtcagtcgtctaaacgtcacccctccgccaagcgcgcgcacc-----aggaggt 590

QY 401 gtgattaccctcgtcgtcgttgcattttgtcttgaccgttactcgcgttcctccaccg 460
    || || || || || || || || || || || || || || || || || || || || || || ||
DB 591 gtggtctctccaaaggcaggtttgaaatcttatctttaaccggtgcttctcctccgga 650

QY 461 cctgcaccacccgggagcagagggtttgacggtgtatctatagccggaggtcaggagcaagtt 520
    || || || || || || || || || || || || || || || || || || || || || || ||
DB 651 ccttcccccacccgggtcaacccggtttaacggttacttagccgggtccagggtcaggtc 710

QY 521 gtaggagggaatgtggtgttcgttaattgcttcggaccggttagtgttgatggctgt 580
    || || || || || || || || || || || || || || || || || || || || || || ||
DB 711 gttggaggtagcgtttagtggccacctcttagccatagggtcgttcattggtgattgctgt 770

QY 581 tctttgcaaacgcagtttatgatgattaccgattgagagaggaagaa 628
    || || || || || || || || || || || || || || || || || || || || || || ||
DB 771 actttctaacgtacttatgagagatgccatggagagagagaa 818

RESULT 6
AAC42265
ID AAC42265 standard; DNA; 1097 BP.
XX
AC AAC42265;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 34898.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
```

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140693.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.


```
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146380.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 18.4%; Score 179.2; DB 21; Length 1192;
Best Local Similarity 58.4%; Pred. No. 2,1e-46;
Matches 354; Conservative 0; Mismatches 243; Indels 9; Gaps 2;
QY 38 gaacccacactgatacaataacatggaacttaacagatctggaagcagacagcctctggtcttctcc 97
Db 158 gaccatgacttcacacccaacacagtggaagccctaactcagactcagagccaagaa 217
QY 98 gcgagaccactcccacgggtgagccaccagctcagccacagcctctggtcttctcc 157
Db 218 gaacgaacagcagagcagcagcagcctgttgaccccggtatccgggtctcag 277
QY 158 ggacgtcgtctccagtggtgctgctcaggttcccaaaacacaccccaaacctccgcagatt 217
Db 278 ggtcgtcgtctcagagtagaccctcctggttccaaagacacacacacagagtcagttgtt 337
QY 218 ataactagagatagtcactaacgttccttagatcacacagcttcttgagagtcacctccggttcg 277
Db 338 gttaccaaagaaagccctaactctctccagagccatgttcttgagattgctacgggagct 397
QY 278 gacatatccgagggcagctctccactcagccactcgtcgcggtgagcgttttgattata 337
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:14:33 ; Search time 56.06 Seconds
(without alignments)
462.792 Million cell updates/sec

Title: us-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETPTGG.....IYMGSGSGGGGGGATRPAPF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392	100.0	270	2 T06118	hypothetical prote
2	706.5	50.8	311	2 F86341	hypothetical prote
3	646.5	46.4	292	2 C71448	hypothetical prote
4	645.5	46.4	302	2 H96792	unknown protein F1
5	629.5	45.2	317	2 D84890	probable AT-hook D
6	625.5	44.9	339	2 T06612	hypothetical prote
7	616	44.3	285	2 E84766	probable AT-hook D
8	605	43.5	324	2 T04576	hypothetical prote
9	581	41.7	265	2 T47898	hypothetical prote
10	572	41.1	310	2 T47695	hypothetical prote
11	554	39.8	257	2 B84860	hypothetical prote
12	417	30.0	206	2 G86279	Fi4L17.27 protein
13	274	19.7	296	2 G84747	AT-hook DNA-bindin
14	255.5	18.4	334	2 T04572	hypothetical prote
15	246.5	17.7	347	2 T06584	probable DNA-bind
16	243	17.5	439	2 T52291	probable DNA-bind
17	239.5	17.2	404	2 T05553	hypothetical prote
18	239	17.2	455	2 T05081	hypothetical prote
19	227	16.3	365	2 T06615	hypothetical prote
20	225	16.2	348	2 T02462	probable AT-hook D
21	218	15.7	345	2 T01348	hypothetical prote
22	211.5	15.2	348	2 T47923	probable DNA-bind
23	162.5	11.7	574	2 A84782	hypothetical prote
24	146	10.5	405	2 T29167	hypothetical prote
25	139	10.0	307	2 T27609	hypothetical prote
26	137	9.8	694	2 T08068	hypothetical glyci
27	135.5	9.7	290	2 T23416	hypothetical prote
28	134.5	9.7	839	2 F75518	hypothetical prote
29	131	9.4	388	2 T29173	hypothetical prote

30 131 9.4 1381 2 E70806 hypothetical glyci
31 129.5 9.3 575 2 S35327 protein kinase agg
32 129.5 9.3 1844 2 T51890 related to Nup98-N
33 129.5 9.3 13288 2 T03099 mucin, submaxillar
34 128.5 9.2 1585 2 T31611 hypothetical prote
35 128 9.2 2468 2 A83412 hypothetical prote
36 125.5 9.0 2038 2 A43742 female sterile hom
37 125 9.0 528 2 D46449 hypothetical prote
38 124.5 8.9 867 2 S57795 probable deoxyribo
39 124 8.9 1655 2 T13998 gene mastermind pr
40 124 8.9 3190 2 T13828 CREB-binding prote
41 122 8.8 615 2 H70589 hypothetical glyci
42 122 8.8 979 2 A35913 regulatory factor
43 121.5 8.7 670 2 T49510 fibroin-3 related
44 121.5 8.7 1231 2 S30185 insulin receptor s
45 121 8.7 418 2 B64924 hypothetical prote

ALIGNMENTS

RESULT 1

T06118
hypothetical protein F23E12.50 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T06118
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Hoheis
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15485
A:Accession: T06118
A:Molecule type: DNA
A:Residues: 1-270 <BEV>
A:Cross-references: EMBL:AL022604; GSPDB:GN00062; ATSP:F23E12.50
A:Experimental source: cultivar Columbia; BAC clone F23E12
C:Genetics:
A:Gene: ATSP:F23E12.50
A:Map position: 4
C:Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 100.0%; Score 1392; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 6.6e-85;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELNRSEADEAKAETPTGGATSSATSSGSRRRPRGRPAGSKNPKPPTIITRDSNV 60

DB 1 MELNRSEADEAKAETPTGGATSSATSSGSRRRPRGRPAGSKNPKPPTIITRDSNV 60

QY 61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIIISGTGAVTNVTIRQPAAPAGGVITLHGR 120

DB 61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIIISGTGAVTNVTIRQPAAPAGGVITLHGR 120

QY 121 FDILSLTGTLPPAPPGAGGLTVYLAGGQVVGNGVAGSLIASPVVLMASAFANAY 180

DB 121 FDILSLTGTLPPAPPGAGGLTVYLAGGQVVGNGVAGSLIASPVVLMASAFANAY 180

QY 181 DRLPPIEETPPRTGVQOQPEASQSSEVTGSGAQACESNLQGGGGGVAFYNLGMN 240

DB 181 DRLPPIEETPPRTGVQOQPEASQSSEVTGSGAQACESNLQGGGGGVAFYNLGMN 240

QY 241 MNPFQSGGDIYMGSGGGGGGGGATRPAPF 270

DB 241 MNPFQSGGDIYMGSGGGGGGGGATRPAPF 270

RESULT 2

F86341
hypothetical protein F9H16.12 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C:Accession: F86341

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

schueller@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES

source

1. 86710
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
173..1890

gene

/gene="F23E12.10"
173..374

exon

/gene="F23E12.10"
173..374

CDS

Join(<173..374,456..919,1089..1610,1678..1890)
/gene="F23E12.10"
/codon_start=1
/product="hypothetical protein (fragment)"
/protein_id="CAA18726.1"
/db_xref="GI:3080407"
/db_xref="SPTREMBL:O65486"

/translation="FNAVAGCFFAVESVLPSSSTSLPNTTSMVLSAVTASV
VSEIGLSEPAKVPDYDFRSPGELPLVLLGALGLVSLALSRCTSSMTSAVDLSNK
VKIATACRASGLGVGVIILVYEVLYMGFQNDILLEKRPVKGLSADLLQLVA
RYGLVMAATLACVQCPLTAVLLFELTDYIRIVLPPLGAVGMSWITSGQSRQPT
RETRTKRKSQEAQVSLTSDDESSTNLCVESSICLDLSNSELPKSFVSEA
MRTPATVMTSLSEALTRMLIEKQSCALIVDPNIFLGLTSLDSIQEFSKARKEG
NRPKIDIFNDICSRSGRCKRVPVTVTPMDMLAAQTIMNKHELSHVAVVSGSIDAPRI
HPVGVLDECIITLRR"
375..455
/gene="F23E12.10"
/number=1
456..919
/gene="F23E12.10"
/number=2
920..1088
/gene="F23E12.10"
/number=2
1089..1610
/gene="F23E12.10"
/number=3
1611..1677
/gene="F23E12.10"
/number=3
1678..1890
/gene="F23E12.10"
/number=4
3124..3241
/gene="F23E12.20"
/number=1
3124..4428
/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

intron

/gene="F23E12.10"
/number=1
456..919
/gene="F23E12.10"
/number=2
920..1088
/gene="F23E12.10"
/number=2
1089..1610
/gene="F23E12.10"
/number=3
1611..1677
/gene="F23E12.10"
/number=3
1678..1890
/gene="F23E12.10"
/number=4
3124..3241
/gene="F23E12.20"
/number=1
3124..4428
/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

exon

/gene="F23E12.10"
/number=1
456..919
/gene="F23E12.10"
/number=2
920..1088
/gene="F23E12.10"
/number=2
1089..1610
/gene="F23E12.10"
/number=3
1611..1677
/gene="F23E12.10"
/number=3
1678..1890
/gene="F23E12.10"
/number=4
3124..3241
/gene="F23E12.20"
/number=1
3124..4428
/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

intron

/gene="F23E12.10"
/number=1
456..919
/gene="F23E12.10"
/number=2
920..1088
/gene="F23E12.10"
/number=2
1089..1610
/gene="F23E12.10"
/number=3
1611..1677
/gene="F23E12.10"
/number=3
1678..1890
/gene="F23E12.10"
/number=4
3124..3241
/gene="F23E12.20"
/number=1
3124..4428
/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

exon

/gene="F23E12.10"
/number=1
456..919
/gene="F23E12.10"
/number=2
920..1088
/gene="F23E12.10"
/number=2
1089..1610
/gene="F23E12.10"
/number=3
1611..1677
/gene="F23E12.10"
/number=3
1678..1890
/gene="F23E12.10"
/number=4
3124..3241
/gene="F23E12.20"
/number=1
3124..4428
/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

intron

/gene="F23E12.10"
/number=1
456..919
/gene="F23E12.10"
/number=2
920..1088
/gene="F23E12.10"
/number=2
1089..1610
/gene="F23E12.10"
/number=3
1611..1677
/gene="F23E12.10"
/number=3
1678..1890
/gene="F23E12.10"
/number=4
3124..3241
/gene="F23E12.20"
/number=1
3124..4428
/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

exon

/gene="F23E12.10"
/number=1
456..919
/gene="F23E12.10"
/number=2
920..1088
/gene="F23E12.10"
/number=2
1089..1610
/gene="F23E12.10"
/number=3
1611..1677
/gene="F23E12.10"
/number=3
1678..1890
/gene="F23E12.10"
/number=4
3124..3241
/gene="F23E12.20"
/number=1
3124..4428
/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

exon

/gene="F23E12.10"
/number=1
456..919
/gene="F23E12.10"
/number=2
920..1088
/gene="F23E12.10"
/number=2
1089..1610
/gene="F23E12.10"
/number=3
1611..1677
/gene="F23E12.10"
/number=3
1678..1890
/gene="F23E12.10"
/number=4
3124..3241
/gene="F23E12.20"
/number=1
3124..4428
/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

gene

/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

CDS

Join(<173..374,456..919,1089..1610,1678..1890)
/gene="F23E12.10"
/codon_start=1
/product="hypothetical protein (fragment)"
/protein_id="CAA18726.1"
/db_xref="GI:3080407"
/db_xref="SPTREMBL:O65486"

intron

/gene="F23E12.10"
/number=1
456..919
/gene="F23E12.10"
/number=2
920..1088
/gene="F23E12.10"
/number=2
1089..1610
/gene="F23E12.10"
/number=3
1611..1677
/gene="F23E12.10"
/number=3
1678..1890
/gene="F23E12.10"
/number=4
3124..3241
/gene="F23E12.20"
/number=1
3124..4428
/gene="F23E12.20"
Join(3124..3241,3323..3465,3595..3696,3945..4107,
4211..4428)
/gene="F23E12.20"
/note="similarity to dihydroflavonol-4-reductase, Gerberta
x sp., PIR2:S35189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translation="MDQAKGVCTVGASGLASVLYKRLLEGVEVIGTVDRDGNKK
LAHLWKEGAKERLRVLYKADLMVEGSDNIMGCGVFHTASPEILPAIEGTLNVLIR
SCRNPSLRKRVLTSSSTVMYLSLKLAQAAMKFSENGIDLTVLPSPFLVPSLP
PDLCSTASDVLLGKGETKRFQWGMGYHIDDVARTHVVFEEHAAQGYICSSNV
ISLEELVSFLSARYPSLPPIPKRFDTPLN"
3242..3322

/gene="F23E12.20"
/number=1
3223..3465
/gene="F23E12.20"
/number=2
3466..3594
/gene="F23E12.20"
/number=2
3595..3696
/gene="F23E12.20"
/number=3
3697..3944
/gene="F23E12.20"
/number=3
3945..4107
/gene="F23E12.20"
/number=4
4108..4210
/gene="F23E12.20"
/number=4
4211..4428
/gene="F23E12.20"
/number=5
complement(Join(4939..5013,5109..5195,5307..5342,
5451..5651,5725..5811,6161..6163))
/gene="F23E12.30"
4939..6163
/gene="F23E12.30"
complement(Join(4939..5013,5109..5195,5307..5342,
5451..5651,5725..5811,6161..6163))
/gene="F23E12.30"
/note="Contains Clathrin adaptor complexes small chain
signature (VYKRYASLYF)
contains EST gb:T44219"
/codon_start=1
/product="clathrin assembly protein API9 homolog"
/protein_id="CAA18728.1"
/db_xref="GI:3080409"
/db_xref="SPTREMBL:O23685"
/translation="MIHFVLLVSRGKVRLTWKVSPYAKQSKRVIRELSGVILNRGP
KLCNFVWRGKVVYKRYASLYFECIDQEDNELEVIHHVYILORYFGSVCELD
LIFPHKAYIILDELLIAGELQESSKTVARIISAQDLVFEVAKESISNIIAQAT
NR"
complement(4939..5013)
/gene="F23E12.30"
/number=1
complement(5014..5108)
/gene="F23E12.30"
/number=1
complement(5109..5195)
/gene="F23E12.30"
/number=2
complement(5196..5306)
/gene="F23E12.30"
/number=2
complement(5307..5342)
/gene="F23E12.30"
/number=3
complement(5343..5450)
/gene="F23E12.30"
/number=3
complement(5451..5651)
/gene="F23E12.30"
/number=4
complement(5652..5724)
/gene="F23E12.30"
/number=4
complement(5725..5811)
/gene="F23E12.30"
/number=5
complement(5812..6160)
/gene="F23E12.30"
/number=5

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:19:08 ; Search time 103.68 Seconds
(without alignments)
450.508 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRFADAKAETPTGG.....IYGMSSGGSGGGGATPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:
1: sp.archaea:
2: sp.bacteria:
3: sp.fungi:
4: sp.human:
5: sp.invertebrate:
6: sp.mammal:
7: sp.mhc:
8: sp.organelle:
9: sp.phage:
10: sp.plant:
11: sp.rodent:
12: sp.virus:
13: sp.vertebrate:
14: sp.unclassified:
15: sp.rv:
16: sp.bacteriap:
17: sp.archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1392	100.0	270	10	065489
2	706.5	50.8	311	10	0957C9
3	646.5	46.4	292	10	023620
4	645.5	46.4	302	10	09C9K7
5	637.5	45.8	315	10	09SR17
6	639.5	45.2	317	10	022130
7	635.5	44.9	339	10	09SZ70
8	616	44.3	325	10	082166
9	605	43.5	324	10	049662
10	581	41.7	265	10	09LZX7
11	572	41.1	310	10	09W2S3
12	554	39.8	257	10	09SUG4
13	499.5	35.9	276	10	09LTA2
14	417	30.0	206	10	09W9R4
15	407	29.2	265	10	0941Y9
16	289.5	20.8	404	10	09LVB0

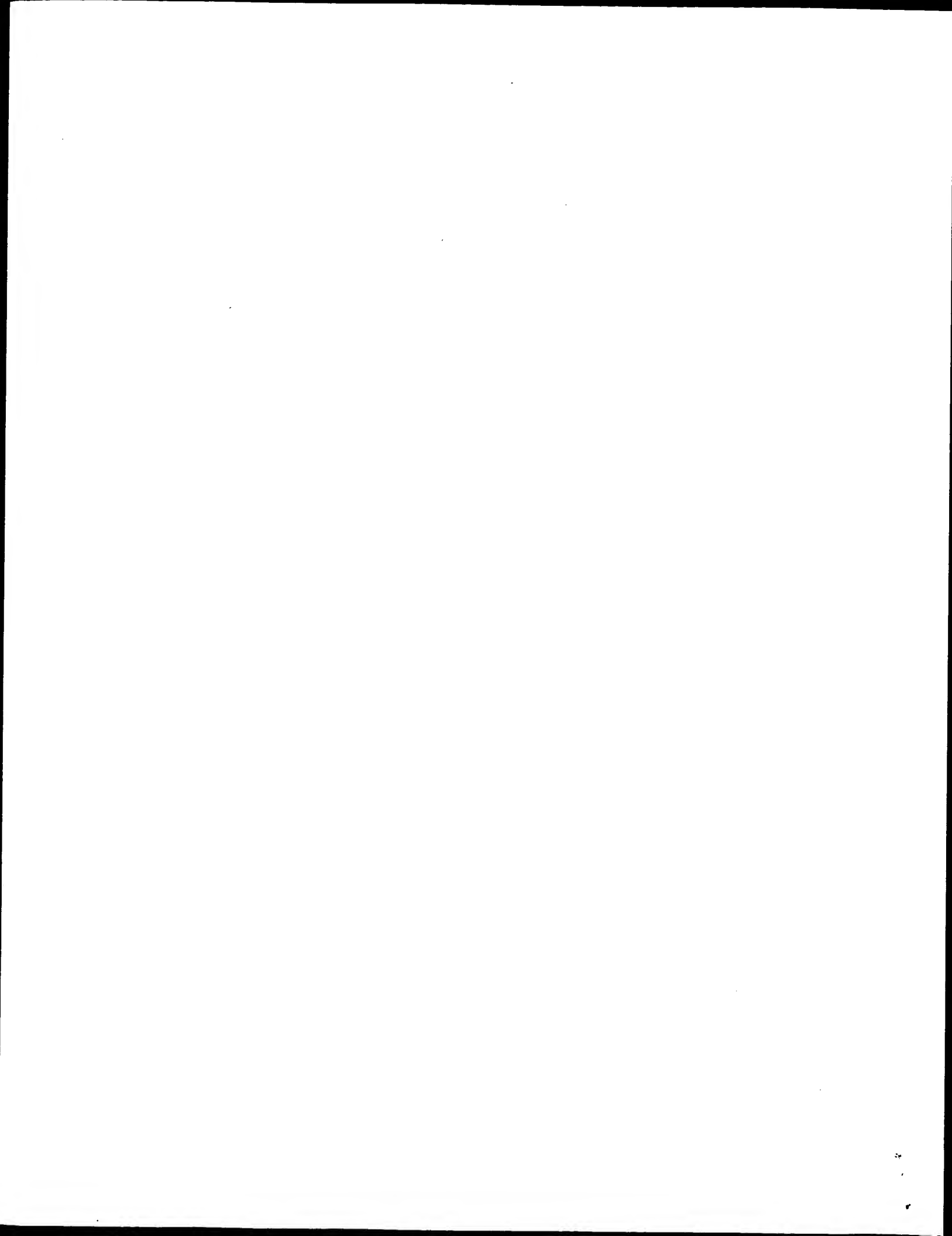
17	289	20.8	771	10	09XHY4
18	283	20.3	405	10	09AYM0
19	282	20.3	351	10	094F52
20	274	19.7	296	10	022812
21	274	19.7	351	10	023142
22	264.5	19.0	419	10	09FHM5
23	259	18.6	826	10	09SH33
24	258.5	18.6	300	10	092RR7
25	255.5	18.4	334	10	049658
26	248.5	17.9	386	10	09FIR1
27	246.5	17.7	347	10	004696
28	243	17.5	439	10	093350
29	243	17.5	439	10	094010
30	242.5	17.4	334	10	004695
31	239.5	17.2	404	10	09SR31
32	239	17.2	455	10	049694
33	227	16.3	365	10	09SZ73
34	225	16.2	348	10	080834
35	221	15.9	418	10	09SR16
36	218	15.7	345	10	081321
37	211.5	15.2	348	10	09M2D3
38	162.5	11.7	574	10	09SJQ2
39	148	10.6	848	5	09V889
40	146	10.5	405	5	Q23057
41	143.5	10.3	818	6	Q9N1P0
42	139	10.0	1468	5	Q9GUB5
43	138.5	9.9	265	5	Q23347
44	137	9.8	694	16	053212
45	136.5	9.8	1610	16	092KQ8

ALIGNMENTS

RESULT 1
065489 PRELIMINARY; PRT; 270 AA.

ID 065489; AC 065489; DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 26.8 KDA PROTEIN.
GN F23E12.50 OR AT4G35390.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Heisel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.F.X.,
RA Schueller C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022604; CAA18730.1; -;
DR EMBL; AL161587; CAB80256.1; -;
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 26774 MW; 4C21487C2719E114 CRC64;

Query Match 100.0%; Score 1392; DB 10; Length 270;
Best Local Similarity 100.0%; Pred. No. 4.7e-86;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

QY 1 MELNRSEADEAKAETTTGGATSSATASGSSGRRPRRGRGAGSKNPKPTTIITRDSNV 60
DB 1 MELNRSEADEAKAETTTGGATSSATASGSSGRRPRRGRGAGSKNPKPTTIITRDSNV 60
QY 61 LRSVLVETSGSDISEAVSYATRRGCGVCIISGTGAVTNTIROPAPAGGCVITLHGR 120
DB 61 LRSVLVETSGSDISEAVSYATRRGCGVCIISGTGAVTNTIROPAPAGGCVITLHGR 120
QY 121 FDLISLTGTALPPPPAGGLTVYLAGGQGVVGGNAGSLIASGPPVLMMAAFANAVY 180
DB 121 FDLISLTGTALPPPPAGGLTVYLAGGQGVVGGNAGSLIASGPPVLMMAAFANAVY 180
QY 181 DRLPIEBEETPPRTTGVQOQPPASOSSEVTGSGAQCENLQGGNGGGVAFYNLGMN 240
DB 181 DRLPIEBEETPPRTTGVQOQPPASOSSEVTGSGAQCENLQGGNGGGVAFYNLGMN 240
QY 241 MNFQSGGDIYMGSGGGGGGATPAF 270
DB 241 MNFQSGGDIYMGSGGGGGGATPAF 270

RESULT 2
ID Q9S7C9 PRELIMINARY; PRT; 311 AA.
AC Q9S7C9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE ESCAROLA
GN ESC OR F9H16.12.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98121113; PubMed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Gielens J., Wambutt R., Weitzenger T., Pohl T.M., Terry N.,
RA Gielens J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA Etian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Rechman S., Ansoorge W., Cooke R., Berger C.,
RA Delsen Y., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chaiwatiz N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana";
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97344; CAB10560.1;
DR EMBL; AL161547; CAB78783.1;
KW Hypothetical protein.
SQ SEQUENCE 292 AA; 29724 MW; 2A948700FA543C4F CRC64;

Query Match 50.88; Score 706.5; DB 10; Length 311;
Best Local Similarity 53.08; Pred. No. 5e-41;
Matches 151; Conservative 25; Mismatches 56; Indels 53; Gaps 6;

QY 7 EADAKAETTTGGATSSATASGSSGRRPRRGRGAGSKNPKPTTIITRDSNVLSHVL 66
DB 59 EDSHNDHQQGRDSDPTSSAPGRPRRGRGAGSKNPKPTTIITRDSNVLSHVL 118
QY 67 EVTSGSDISEAVSYATRRGCGVCIISGTGAVTNTIROPAPAGGCVITLHGR 119
DB 119 EVSPGADIVESVSYATRRGCGVCIISGTGAVTNTIROPAPAGGCVITLHGR 178
QY 120 RFDLSLTGTALPPPPAGGLTVYLAGGQGVVGGNAGSLIASGPPVLMMAAFANAV 179

```

```

DB 179 RFEILSLGTALPPPPAGGLTVYLAGGQGVVGGNAGSLIASGPPVLMMAAFANAV 238
QY 180 YDLPIEBEETPPRTTGVQOQPPASOSSEVTGSGAQCENLQGGNG 228
DB 239 FERLPIEBEETPPRTTGVQOQPPASOSSEVTGSGAQCENLQGGNG 287
QY 229 GGVAFYNLGMNMFQSGGDIYMGSGS---GGGGGATPAF 270
DB 288 GG-----YGFSGDPLHGLWGAGCTPSRPF 311

RESULT 3
ID Q23620 PRELIMINARY; PRT; 292 AA.
AC Q23620;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHETICAL 29.7 KDA PROTEIN.
GN AT4G17800.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98121113; PubMed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Gielens J., Wambutt R., Weitzenger T., Pohl T.M., Terry N.,
RA Gielens J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA Etian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Rechman S., Ansoorge W., Cooke R., Berger C.,
RA Delsen Y., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chaiwatiz N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana";
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97344; CAB10560.1;
DR EMBL; AL161547; CAB78783.1;
KW Hypothetical protein.
SQ SEQUENCE 292 AA; 29724 MW; 2A948700FA543C4F CRC64;

Query Match 46.48; Score 646.5; DB 10; Length 292;
Best Local Similarity 52.78; Pred. No. 5.9e-37;
Matches 136; Conservative 32; Mismatches 47; Indels 43; Gaps 8;

QY 19 GGATSSATASGSSS-----GRRPRRGRGAGSKNPKPTTIITRDSNVLSHVLTSQSD 73
DB 62 GGSAGSSGGGGGGGGGVDVGRPRRGRGAGSKNPKPTTIITRDSNVLSHVLTSQSD 121
QY 74 ISEAVSYATRRGCGVCIISGTGAVTNTIROPAPAGGCVITLHGRDILSLGTALPP 133
DB 122 VEDCVATARRRQRCICVLSGTVNTVSIQPS--AAGAVTLOGTTEILSLSGSFLPP 179
QY 134 PAPPAGGAGTIVYLAGGQGVVGGNAGSLIASGPPVLMMAAFANAVYORLPIEBEETPP 193
DB 180 PAPPAGTSLTIFLAGGQGVVGGVGTAGPVIVIAAFTNVAERLPLEEDE----- 235
QY 194 RTTGVOQOQPPASOSSEVTGSGAQCENLQGGN-----GGGVAFYNLGMN-NF 244

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:14:33 ; Search time 56.06 Seconds
(without alignments)
462.792 Million cells

Title: US-09-823-676-2

Perfect score:

Sequence: 1 MELNRSEADEAKAETPTGG.....IYGMSSGGSGGGGATRP AF 270

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 9
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

100% processing. Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:★

1: p1r1:"

2: p1r2:

3: p1r3:

4: pir4:v

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1392	100.0	270	2	T06118	hypothetical prote	
2	706.5	50.8	311	2	F86341	hypothetical prote	
3	646.5	46.4	292	2	C71448	hypothetical prote	
4	645.5	46.4	302	2	H96792	unknown protein F1	
5	629.5	45.2	317	2	D84890	probable AT-hook D	
6	625.5	44.9	339	2	T06612	hypothetical prote	
7	616	44.3	285	2	E84766	probable AT-hook D	
8	605	43.5	324	2	T04576	hypothetical prote	
9	581	41.1	365	2	T47898	hypothetical prote	
10	572	41.1	310	2	T47695	hypothetical prote	
11	554	39.8	257	2	B84860	hypothetical prote	
12	417	30.0	206	2	G86279	hypothetical prote	
13	274	19.7	296	2	G84747	F14L17.27 protein	
14	255.5	18.4	334	2	T04572	AT-hook DNA-bindin	
15	246.5	17.7	347	2	T06584	hypothetical prote	
16	243	17.5	439	2	T52291	probable DNA-bind	
17	239.5	17.2	404	2	T05553	probable DNA-bind	
18	239	17.2	455	2	T05081	hypothetical prote	
19	227	16.3	365	2	T06615	hypothetical prote	
20	225	16.2	348	2	T04462	probable AT-hook D	
21	218	15.7	345	2	T01348	hypothetical prote	
22	211.5	15.2	348	2	T47923	probable DNA-bind	
23	162.5	11.7	574	2	A84782	hypothetical prote	
24	146	10.5	405	2	T29167	hypothetical prote	
25	139	10.0	694	2	T27609	hypothetical prote	
26	137	9.8	694	2	F70868	hypothetical glyci	
27	135.5	9.7	290	2	T23416	hypothetical prote	
28	134.5	9.7	839	2	F75518	hypothetical prote	
29	131	9.4	388	2	T29173	hypothetical prote	

```
Query Match      100.0%; Score 1392; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 6.6e-85;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

[illegible]

RESULTS

RESOL 2
F86341

hypothetical protein F9H16.12 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C; Accession: F86341

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Thu Jul 11 11:00:58 2002

```
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 18.4%; Score 178.8; DB 21; Length 1190;
Best Local Similarity 63.2%; Pred. No. 2.8e+46;
Matches 311; Conservative 0; Mismatches 172; Indels 9; Gaps 2;

QY 152 tctccggagcgtctccacgtgtgctcctgcagggtcccaaaacaaacccaacacccg 211
    || || || || || || || || || || || || || || || || || || || ||
DB 270 tctacgggtcgtcgtcctagaggtagacctcctggtcccaagaacaaacaaagtcga 329
    || || || || || || || || || || || || || || || || || || || ||
QY 212 acgattataactagatagctcctaaacgtccttagatcacacgttcttgaagtcacctcc 271
    || || || || || || || || || || || || || || || || || || || ||
DB 330 gttgtgtttaccaaagaagccctaaactctccagagccatgttctttagattgctacg 389
    || || || || || || || || || || || || || || || || || || || ||
QY 272 ggttcggacatatccgagggcagctctccacctacgcccactcgtcgcggtcgcggtttgc 331
    || || || || || || || || || || || || || || || || || || || ||
DB 390 ggaagtacgtggcggaagccttaaacgcctttgctcgttagacgcggcggttttcg 449
    || || || || || || || || || || || || || || || || || || || ||
QY 332 attataagcggcagcgggtggtcgtcactaaogtcaacgatacggcaacctgcggtccggct 391
    || || || || || || || || || || || || || || || || || || || ||
DB 450 gtgctgagcggtagtgggtttgttactaatgttactctcgtcgtcagcgtcgtgcacg 505
    || || || || || || || || || || || || || || || || || || || ||
QY 392 ggtgaggtgtgattaccctgcagtggttggttgacattttgtcttgaccggtactgcg 451
    || || || || || || || || || || || || || || || || || || || ||
DB 506 --cgttgaggtgttagtttacgtgttcaggttgatcttgcattctgtctatgtgtgggctttt 563
    || || || || || || || || || || || || || || || || || || || ||
QY 452 ctccaccgcctg---caccacgggagcagaggttttgacgtgtatctatcagcggaggt 508
    || || || || || || || || || || || || || || || || || || || ||
DB 564 ctccctacgtcgtcctcctcgtcgcagcgcgtgttttaaccatttacttagctggagct 623
    || || || || || || || || || || || || || || || || || || || ||
QY 509 caaggacaagtttagaggggaatgtggtcgttcgttcgttaattccttcggaccggttagtg 568
    || || || || || || || || || || || || || || || || || || || ||
DB 624 caaggccaagttgtggaggtgaggtgtggtggttcgtggttcctctgaccggttatt 683
    || || || || || || || || || || || || || || || || || || || ||
QY 569 ttgatgctgctcttttgcacacgagtttatgataggttacccgttagcaggaagaa 628
    || || || || || || || || || || || || || || || || || || || ||
DB 684 gtgatagctgctacgttttgcattgcaacttatgagaggttacogattgaggaagaa 743
    || || || || || || || || || || || || || || || || || || || ||
QY 629 accccaccgcg 640
    || || || ||
DB 744 cagcaagagcag 755

RESULT 9
AAS96327
ID AAS96327 standard; cDNA; 1251 BP.
XX AC AAS96327;
XX AC AAS96327;
XX AC AAS96327;
DT 26-FEB-2002 (first entry)
XX DE Arabidopsis cDNA encoding cell cycle protein CCP27.
XX DE Arabidopsis thaliana.
KW Cell cycle protein; CCP; ss; cell cycle regulation; herbicide;
KW plant growth regulator; plant development; abiotic stress; biotic stress;
KW nutrient deprivation; pathogen attack; crop yield.
XX OS Arabidopsis thaliana.
XX PN WO200185946-A2.
XX PD 15-NOV-2001.
XX PF 14-MAY-2001; 2001WO-IB01307.
```


PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136382.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0148171.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148319.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148565.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148684.
PR	17-JUN-1999;	99US-0139432.	PR	13-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149922.
PR	18-JUN-1999;	99US-0139459.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139461.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139462.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139463.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	99US-0139817.	PR	30-AUG-1999;	99US-0151303.
PR	22-JUN-1999;	99US-0139899.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140353.	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	99US-0140354.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0140695.	PR	10-SEP-1999;	99US-0153070.
PR	24-JUN-1999;	99US-0140823.	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	99US-0140991.	PR	15-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0141842.	PR	20-SEP-1999;	99US-0154779.
PR	01-JUL-1999;	99US-0142154.	PR	22-SEP-1999;	99US-0155139.
PR	02-JUL-1999;	99US-0142055.	PR	23-SEP-1999;	99US-0155486.
PR	06-JUL-1999;	99US-0142390.	PR	24-SEP-1999;	99US-0155659.
PR	08-JUL-1999;	99US-0142803.	PR	28-SEP-1999;	99US-0156458.
PR	09-JUL-1999;	99US-0142920.	PR	29-SEP-1999;	99US-0156596.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0157117.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157753.
PR	14-JUL-1999;	99US-0144005.	PR	05-OCT-1999;	99US-0157865.
PR	15-JUL-1999;	99US-0144085.	PR	06-OCT-1999;	99US-0158232.
PR	16-JUL-1999;	99US-0144086.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144325.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144331.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144334.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144335.	PR	13-OCT-1999;	99US-0159329.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144684.	PR	14-OCT-1999;	99US-0159637.
PR	21-JUL-1999;	99US-0144814.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0145086.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160741.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	23-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	26-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	23-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.

us-09-823-676-1.rng

Thu Jul 11 11:00:58 2002

PN	EPI033405-A2.	PR	01-JUL-1999;	99US-0142154.
XX		PR	02-JUL-1999;	99US-0142055.
FD	06-SEP-2000.	PR	06-JUL-1999;	99US-0142390.
FF		PR	08-JUL-1999;	99US-0142803.
FF		PR	09-JUL-1999;	99US-0142920.
XX	25-FEB-2000; 2000EP-0301439.	PR	12-JUL-1999;	99US-0142977.
XX		PR	13-JUL-1999;	99US-0143542.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	PR	22-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145087.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	PR	30-AUG-1999;	99US-0151303.
PR	22-JUN-1999;	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	PR	15-SEP-1999;	99US-0154019.
PR	30-JUN-1999;	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	PR		

us-09-823-676-1.l.rng

Thu Jul 11 11:00:58 2002

```
Db 604 ttgcgttttatgcgaacaggtgtgctgttttcgagcgttaacacttcgctcagcctgattcat 663
QY 387 cggctgtgtgaggtgtgattaccctgcgtatgcgtgtttgacattttgtctttgaccggtta 446
Db 664 c-----tgggtgtacattgacctatgaggtgcgttttggatatttcacatactctgaa 717
QY 447 ctgcgtctc-----cacgcctgcaccaccggtgagcagggaggtttgacgtgtgtatc 497
Db 718 cattcatgcctagtactcagacgggacacagacagcagacagcgggatgagcgtgtcgc 777
QY 498 tagcggaggtcgaagacaagatttagaggaagatgtgcgtgttcgttaattcttcgg 557
Db 778 ttgctagcctgatgagcgtgtagtagtgggtgtgtgtgctggtgctgctgctggttcgagcca 837
QY 558 gaccggtagtgtgtgctgtctttt 585
Db 838 ctctctaatcaggtgttagaacttt 865

RESULT 14
AAC38928
ID AAC38928 standard; DNA; 1387 BP.
XX
AC AAC38928;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 22725.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EF1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 19:59:17 ; Search time 63.21 Seconds
(without alignments)
3784.955 Million cell updates/sec

Title: US-09-823-676-1
Perfect score: 974
Sequence: 1 cccccgcacctgcctctaca.....cgaatgcgaattatagtt 974

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCrus_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	44.6	4.6	7218	1	US-08-232-463-14
C 2	33.8	3.5	8868	3	US-08-815-809-7
C 3	33.8	3.5	19877	2	US-08-816-155B-8
C 4	33.8	3.5	19877	3	US-09-079-587-8
C 5	33.6	3.4	2793	1	US-08-209-747-1
C 6	33.6	3.4	2793	1	US-08-458-298-1
C 7	33.6	3.4	9468	1	US-08-325-547-10
C 8	33.4	3.4	1100	3	US-08-642-807A-32
C 9	31.8	3.3	4403765	4	US-09-103-840A-2
C 10	31.6	3.2	5433	3	US-08-929-329-1
C 11	31.6	3.2	22671	4	US-08-976-329-14
C 12	31.4	3.2	765	4	US-09-199-637A-324
C 13	31.4	3.2	1071	4	US-09-199-637A-322
C 14	31.4	3.2	1590	4	US-09-199-637A-320
C 15	31.4	3.2	42235	4	US-09-199-637A-1
C 16	31.2	3.2	720	4	US-08-998-416-859
C 17	31.2	3.2	1028	4	US-08-118-200-1
C 18	31.2	3.2	1028	4	US-08-458-745-1
C 19	31.2	3.2	6085	4	US-09-029-603-4
C 20	31.2	3.2	80246	4	US-09-078-294-4
C 21	31	3.2	2790	2	US-08-718-661-1
C 22	31	3.2	6306	1	US-08-466-390-3
C 23	31	3.2	6306	1	US-08-470-950-3
C 24	31	3.2	6306	1	US-08-467-781-3
C 25	31	3.2	6306	1	US-08-195-487-3
C 26	31	3.2	6306	2	US-08-483-924-3
C 27	31	3.2	6306	5	PCT-US93-06160-3

28	31	3.2	7218	1	US-08-232-463-14
29	30.8	3.2	407	4	US-09-056-556-173
30	30.8	3.2	1683	1	US-07-945-283-3
C 31	30.8	3.2	1831	6	5215881-1
C 32	30.8	3.2	1831	6	5215881-3
C 33	30.8	3.2	8438	1	US-07-945-283-1
C 34	30.8	3.2	3915	2	US-08-485-139-6
C 35	30.6	3.1	3915	3	US-08-750-357-6
C 36	30.6	3.1	4137	3	US-08-750-357-7
C 37	30.6	3.1	31571	1	US-08-323-443B-1
C 38	30.6	3.1	5526	3	US-08-658-136-2
C 39	30.6	3.1	53577	3	US-08-658-136-1
C 40	30.4	3.1	396	2	US-08-244-205-14
C 41	30.4	3.1	396	5	PCT-US92-10284-14
C 42	30.4	3.1	3541	4	US-08-976-259-80
C 43	30.4	3.1	6530	2	US-08-145-930-1
C 44	30.4	3.1	6530	3	US-08-458-240-1
C 45	30.4	3.1	6530	5	PCT-US93-03993-1

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: ptzgt-F1s
US-08-232-463-14

Sequence 14, Appl
Sequence 173, App
Sequence 3, Appli
Patent No. 5215881
Patent No. 5215881
Sequence 1, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 7, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 14, Appli
Sequence 14, Appli
Sequence 80, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli

```

Db      7979 ttatctcaagatatgaagaagataatc 8007

RESULT      3
US-08-816-155B-8
; Sequence 8, Application US/08816155B
; Patent No. 5990091
; GENERAL INFORMATION:
; APPLICANT: TARTAGLIA, JAMES
; APPLICANT: COX, WILLIAM I.
; APPLICANT: GETTIG, RUSSELL R.
; APPLICANT: MARTINEZ, HECTOR
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: PINCUS, STEVEN E.
; TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
; TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentLin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,155B
;

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: KOWALSKI, THOMAS J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454310-2990
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-588-0800
; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19877 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-816-155B-8

Query Match 3.5%; Score 33.8; DB 2; Length 19877;
Best Local Similarity 51.7%; Pred. No. 10;
Matches 77; Conservative 72; Indels 0; Gaps 0

Qy 483 gttgacaggttatctagccgagctcaagagacaagttaggaggggaatgtgctgggt 542
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7909 GTAAATCGGTGTTATTATTCCTAGAGATCTATAATTATATATAGATTGTTGTTT 7968
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	7969	TTTAAATTGCTACATACCGGTCTTCATAAAAGATGATTATTAGATAAAAGATAGTCTATC	8028
Qy	603	atagsttaccgattgaagaggaagaacc	631
Db	8029	TTATCTACAAGATATGAAAGAAGATAATC	8057

RESULT⁴

US-09-079-587-8

; Sequence 8, Application US/09079587

; Patent No. 6130066

; GENERAL INFORMATION:

; APPLICANT: TARTAGLIA, JAMES

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,547
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/759,570
FILING DATE: 12-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: Wetherell, Jr., Ph.D., John R.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: PD-2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 9468 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: P
FEATURE:
NAME/KEY: CDS
LOCATION: join(628..1977, 1869..5240, 5236..5988, 5992
LOCATION: ..6222, 6262..8824, 6709..6912)
US-08-325-547-10

```

; OTHER INFORMATION: /product= "N. clavipes minor
;
; OTHER INFORMATION: ampullate silk protein"
US-08-458-298-1

```

	Query Match	3.48;	Score 33.6;	DB 1;	Length 2793;
	Best Local Similarity	53.9%;	Pred. No. 4.4;	59;	Indels 0;
	Matches 69;	Conservative 0;	Mismatches 59;	Indels 0;	Gaps 0;
QY	268	ctccgggtcggacatctccgaggcagctccacactacgccactctgcggctgcggcgt	327		
Db	1388	CACCGTACCACCTTGACCCAGGTAGCTCCAGTCTGTGACCCGCGAGGGCTCTTGTC	1329		
QY	328	tgtcattataagcgacgggtgcggtcactaaagtcacgatacgcgcaacttcgcggctcc	387		
Db	1328	CAGCACCGCAGCGGCTCTGCTCCAGCAGCAGCACCTACCGCTAACCCAGCGGCTCC	1269		
QY	388	ggctgggtg	395		
Db	1268	AGTCTCTG	1261		

```

RESULT      7
US-08-325-547-10/c
; Sequence 10, Application US/08325547
; Patent No. 5736378
;
; GENERAL INFORMATION:
;
; APPLICANT: Elder, John R.
;
; APPLICANT: Talbott, Randy L.
;
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE AND OPEN READING
;
; TITLE OF INVENTION: FRAMES OF FELINE IMMUNODEFICIENCY VIRUS (FIV)
;
; NUMBER OF SEQUENCES: 10
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESS: Spensley Horn Jubas & Lubitz
;
; STREET: 1880 Century Park East, Suite 500
;
; CITY: Los Angeles
;
; STATE: California
;
; COUNTRY: USA
;
; ZIP: 90067

```

[illegible]

```

RESULT      8
US-08-642-807A-32/c
Sequence 32, Application US/08642807A
Patent No. 6087097
GENERAL INFORMATION:
APPLICANT: Persing, D. H.
TITLE OF INVENTION: PCR DETECTION OF BORRELIA
TITLE OF INVENTION: BURGDOFFER
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
STREET: P.O. Box 2938
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0

```

```

Query Match          3.3%; Score 31.8; DB 4: Length 4403765;
Best Local Similarity 56.1%; Pred. No. 52;
Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```

	Query Match	3.2%	Score 31.6;	DB 3;	Length 5433;
	Best Local Similarity	52.2%;	pred. No. 23;		
Matches	70; Conservative	0;	Mismatches	64; Indels	Gaps 0;
QY	511	aggacaagttgagaggggaatgctgcgtgttcgttaattgcttcggggaccggtagtgtt	570		
Ddb	3442	AAGAAAATTTCGAAGAACACAAAAAGGTGCTGGCGCAAAAAAAGCAGAAAGCAAAGGAA	3501		
QY	571	gatggctgcctcctttgcaacgcagtttatgataggttaccgattgaaggggaagaac	630		
Ddb	3502	AAAGCTGAAGCTCTAAAAAAGCACAGGAGGAAAGAAACGAATTTGAAGCTGAAAGAA	3561		
QY	631	cccaccgcccgagaa	644		
Ddb	3562	AGCAGAGGAGGAAA	3575		

Thu Jul 11 11:00:59 2002

us-09-823-676-1.rni

Page 8

Search completed: July 10, 2002, 21:34:24
Job time: 5707 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 19:01:07 ; Search time 1926.27 Seconds
(without alignments)
6824.606 Million cell updates/sec

Title: US-09-823-676-1

Perfect score: 974

Sequence: 1 cccccgacctgctctaca.....cgaaatgcgaatattaggtt 974

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estro: *
7: em_estro: *
8: em_hic: *
9: gb_estl: *
10: gb_est2: *
11: gb_hic: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vrt: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	208.6	21.4	759	10	BG581882 EST483618
2	206	21.1	597	9	AW585780 EST317403
3	198	20.3	736	10	BG583651 EST485403
4	196.6	20.2	876	12	BH550632 EST485403
5	192.4	19.8	688	9	AW774484 EST333635
6	192.4	19.8	690	9	AW574000 EST316591
7	192.4	19.8	765	10	BG647144 EST508763
8	190.8	19.6	671	10	BG587531 EST489303
9	187.6	19.3	449	9	BE203784 EST396460
10	180.4	18.5	782	12	BH419879 EST489303
11	179.2	18.4	534	10	BG134451 EST467343
12	178.4	18.3	765	12	BH481983 EST467343
13	177.6	18.2	488	9	AL366947 EST335147
14	174	17.9	582	9	AW776082 EST335147
15	173.4	17.8	747	12	BH566718 EST335147
16	166.8	17.1	745	12	BH596283 EST335147
17	165.4	17.0	618	9	AW980581 EST335147

18	163.8	16.8	715	10	BM110212
19	158.8	16.3	531	9	AW560824
20	158.8	16.3	555	10	BI321563
21	158.6	16.3	395	10	BF067277
22	158.2	16.2	652	10	BE999313
23	157.8	16.2	762	12	BH459056
24	157.6	16.2	750	10	BG583687
25	156.8	16.1	591	10	BF254863
26	150	15.4	783	9	AW349284
27	148.8	15.3	451	12	BH419964
28	144.2	14.8	583	10	BI701170
29	143.8	14.8	792	12	BH513264
30	140	14.4	739	9	AW350603
31	139.8	14.4	394	9	AW774872
32	138.4	14.2	628	9	AW720668
33	138.2	14.2	449	10	BF004270
34	137.2	14.1	542	12	BH550963
35	137	14.1	512	10	BE998262
36	136.4	14.0	417	10	BI419871
37	129	13.2	737	10	BG583795
38	126.2	13.0	309	12	BH173434
39	125.8	12.9	638	9	AV544413
40	125.6	12.9	521	10	BG881387
41	125.2	12.9	456	10	T43108
42	123.2	12.6	448	9	AV422634
43	117.4	12.1	673	12	BH430946
44	116.2	11.9	458	9	AI494847
45	115.2	11.8	335	9	BE020549

ALIGNMENTS

RESULT 1

LOCUS	BG581882	759 bp	mRNA	linear	EST 11-APR-2001
DEFINITION	EST483618	GVN Medicago truncatula	cdna clone	pgvn-66k1 5' end,	mrna
ACCESSION	BG581882				
VERSION	BG581882.1	GI:13596946			
KEYWORDS	EST				
SOURCE	barrel medic.				
ORGANISM	Medicago truncatula				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.				
AUTHORS	1 (bases 1 to 759) Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town ,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.				
TITLE	ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula, 2001				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Carroll P. Vance Department of Agronomy and Plant Genetics University of Minnesota 411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA Tel: 612 625 5715 Fax: 651-649-5058 Email: vance004@maroon.tc.umn.edu				

University of Minnesota name: M82944e TIGR sequence name:
MTCDO61TK More information is available at: <http://www.medicago.org>
Seq primer: SKmod (CTA gaa cta gta gat CC).
Location/Qualifiers
1. 759
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pgvn-66k1"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"

691	CGAAGCCACCGGTGATTGTGACAAGAGATAGCCCAACGTCGTTAGATCTCATGTTCTTG	632
261	aagtcacotccggttcgacatatccgaggcagctccacetaacgcccactctgcgcgct	320
631	AACTCTCATCTGGAGCCGACATAATTGAGAGGTCACAAATTACGCTCGCCGAAGGGA	572
321	gogcggttgcaattataagcgcacggtgcggtcacataacgatacagatacggcaactcg	380
571	GAGGTGTCCTCAATTCTCAGTGTAAACGGCGCGTGGCTAACCTACTCTCCACAGCGGG	512
381	cggtctcggctgg-----tgaggtgtgattaccctgc	413
511	TGACGACTCATGGGAACAATGTTGGAACTGAAAGCGGAGCTGGAGGATGTGTACTTTAC	452
414	atggtcggtttaacattttgtcttgaccggtactgcgtccaccgcgctgcaccacgg	473
451	ATGGAAGGTTTGAGATCTTTCATCACCATTGGTACGGTGTCTCCGCGCCGCCGCCGG	392
474	gagcaggaggttgacggtgtactcagccgaggtcaaggacaagtgttagagggaatg	533
391	GATGCGGTGGTTTATCTATCTTTGTTGCTGGTGAAACAAGTCCGATGATCGGAGGAAGAG	332
534	tggtcgtgttcgtaattgcttogggacoggtadgtgtgctgctgtctcttttgcaaacg	593
331	TGCTGGCTCCGCTTGTGGCTTCGTCCGCTGATCTGATGGCTGCATCGTTCTCCAACG	272
594	cagtttatgataggttacccgatigaagagaagaa	628
271	CAGACTTTTGAAGGCTTCCGCTTTGAAGATGAAGGA	237

RESULT	5
AW774484	
LOCUS	linear
	688 bp
	mRNA
	EST 07-SEP-2000
	AW774484
	cDNA clone pKV3-2p11, mRNA
	truncated
	virusMedic90
	ncm220235

ACCESSION AW774484
VERSION AW774484.1 GI:7718401
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE
1 (bases 1 to 688)
VandenBosch, K., Hurt, J., Moore, J., Beremand, P., Peng, H., Ellis, L., Town, C. D., Bowman, C. L., Craven, M. B., Hansen, T. S., Holt, I. E. and Fraser, C. M.

TITLE	JOURNAL	COMMENT
Fisher, C.M. ESTs from roots of Medicago truncatula after Rhizobium inoculation	Unpublished (1999)	Contact: VandenBosch K Department of Biology

Department of Biology
Texas A&M University
College Station, TX 77843-3258, USA
Tel.: 409 845 7707
Fax: 409 845 2891

Fax: 409 845 2891
Email: kate@mail.bio.tamu.edu
Texas A&M EST name: T257903e
TIGR sequence name: MTEBA90TK
http://www.tamug.edu/~kate/

FEATURES

More information is available at .
<http://chrystie.tamu.edu/medicago>
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
Location/Qualifiers

```
source
1..688
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
```

```

/obj_...
/clone="pkV3-22p11"
/clone_lib="KV3"
/tissue_type="Seedling roots"
/dev stages="3 days post-inoculation with Sinorhizobium

```

```

/dev_stage= 3 days post incubation.
melliloti#

```

```

/lab_host="E. coli strain XL0LR"
/notes="Vector: pBluescript SK -; Site.1: EcoRI; Site.2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XL0LR cells."
177 a 145 c 182 g 184 t

```

[illegible]

202 ATACAAAGAGGAGGAGTGCACACACACTTTAGGCGTCACACATCCTTTGAAGTAAGTATGATGGTGT 261
Db
278 gacatatacgaagcagctccactacgcacactcgctcgcgctcgcgcttgcattatata 337
Qy
262 GATGTGTTTACTCGTGTGGCTACTTATGCAAGAAGGCTCAAGAGGAGATCGTGTCTCT 321
Db
338 agcgcgcagcggtcggfcaactaacgtcacatcacgcgaacacctgcgctcccgcgctgggtaga 397
Qy

Db	322	AGTGGGAGTGAACAGCTCTATACGTGACCTTGCAGACGCCAG-----CTGGCGCAGGA	375
Qy	398	ggtagattacctgcgatgctcgatttgacattttctttgacgggtacgcgcgtttcca	457
Db	376	TCTGTAGTGACTTTACATGGAAGGTTTGAATACTTCTTTTGTCCGGATCGTTTCTACCA	435
Qy	458	ccgcctgcacacacgcggagagcagggtttgacggttgtagctatcagccggagggttcaaggacaa	517

Db	436	CCACGAGCTCTCCAGGAGCTACAAAGTTTGAATGTGTTCCCTTGGTGAGGCCAAGTCAA	495
Qy	518	gttgaggagggaatgtggtctgtaattgcttcggacacgtagtttggatgagct	577
Db	496	GTTGTGGGAGGAATGTTGTGGTCCTTGGTGGTCTTGGACCGTATTGTCATGCT	555
Qy	578	gctcttttgcacacgcagtttatgataggtagtcagattgaagaggaggaacccacgcg	637

[illegible]

sequence.
 ACCESSION AW574000
 VERSION AW574000.1 GI:7238733
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM *Medicago truncatula*
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Magnoliophyta; eudicotyledons; core eudicots; Spermatophyta; Magnoliopsida; trifoliales

REFERENCE
AUTHORS
Rosidae; eusoids I; Fabales; Fabaceae; Papilionoideae; Trifoliales
Medicago.
1 (bases 1 to 690)
Fedorova, M., pietson, B.L., Samac, D.A., Vance, C.P., Gantt, J.S., Pen-
H. Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S.,
Holt, I. E. and Fraser, C.M.
ESTs from one month old nitrogen-fixing root nodules of Medicago
truncatula

truncatula
JOURNAL
COMMENT
Unpublished (2000)
Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota


```

, C. and Tanksley, S.
Generation of ESTs from tomato crown gall tissue
Unpublished (2001)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1. .534
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="tomato crown gall"
/tissue_type="crown gall"
/dev_stage="crown galls from full-grown plants (8 wks old
)"
/lab_host="SOLR"
/note="vector: pbluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Four wk old greenhouse plants were stab inoculated
on stem with Agrobacterium tumefaciens C58 (Dr. T.J. Burr,
Cornell U.). Galls were allowed to develop for another 4
wks, when gall tissue was frozen in liquid nitrogen."
BASE COUNT      119 a 137 c 138 g 140 t
ORIGIN

Query Match      18.4%; Score 179.2; DB 10; Length 534;
Best Local Similarity 62.7%; Pred. No. 1.3e-38;
Matches 302; Conservative 0; Mismatches 168; Indels 12; Gaps 1;

QY 146 ggctcttcctccgaagcgtctccagtggtgtctctcaggttccaaacaaacccaaa 205
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 GCGGAGCTCCCAATCGCGCTCGAGCGCTCGCGCTGGATCTAAATAAGCCCAAG 96

QY 206 cttccgagattataactagatagtccttaacgtcttagatcacacgttcttgaagtc 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 CCTCAATCATCGTGACGAGAGATACGGCTAACGCACCTCCGATCTCACGTCTTGAAGTT 156

QY 266 acctcgggttcggacatatccgagcgagtcctccacctacgcccactcgtcggcgctgcgc 325
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 157 TCGACCGATGTTGATATCATGGAAAGTATCTCCCAATTACGCAAGCGGAGAGGAGGT 216

QY 326 gttgcattataagcgcaaggtgcggtcactaacgtcacgatacagcgcaacctgcggt 385
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 217 GTTTGATTTCTAGTGTAGCGGCACAGTTACCAACCTTCTGTCAGCCCTGC--- 272

QY 386 cggcgctggtgagtgattaccctgcgtgctggtttgacatttgtcttgaccggt 445
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 -----TGCAGTGTAGTACACTCCAGGACGTTTCGAATACTTAGCCCTCAGGT 324

QY 446 actggtctccaccgctgcaccacgggagcaggaggtttgacggtgtatctagccgga 505
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 ACGGTGCTTCTCGCCTGACCCGCCCTCCAGTCCAGTGGATCTCTATATTTTATCAGGT 384

QY 506 ggtcaagacaaagtttagagggaagtgtggtggttcgttaattgtctcggaaccggtta 565
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 GGACAAGACAAAGTGGTGGAGGATCCGTTGTAGGGGCTTTGATCGCATCAGGTCCAGTC 444

QY 566 gtgttgatggtctctcttttgcaaacgacagtttatgataggttaccgattgaagagaa 625
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 445 GTCTTAATGGCTGCTCTTTTGCTAAATGCTGTATTTGAACGACTTCCCTTGGAGGAAGAT 504

QY 626 ga 627
|||
Db 505 GA 506

RESULT 12
BH481983
LOCUS
DEFINITION BOGYI73TR BOGY Brassica oleracea genomic clone BOGYI73, DNA
765 bp DNA linear GSS 13-DEC-2001
BH481983

```

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

sequence.
BH481983
BH481983.1 GI:17690087
GSS.
Brassica oleracea.
Brassica oleracea

REFERENCE AUTHORS TITLE JOURNAL COMMENT

1 (bases 1 to 765)
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOGYI73TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES source

1..765
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGYI73"
/clone_lib="BOGY"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 181 a 161 c 215 g 208 t
ORIGIN

Query Match 18.3%; Score 178.4; DB 12; Length 765;
Best Local Similarity 63.2%; Pred. No. 2.5e-38;
Matches 306; Conservative 0; Mismatches 151; Indels 27; Gaps 1;

QY 172 tggctcctcctcaggttccaaaacaaacccacccacgacgattataactagatagatag 231
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 TGTTCTCTCTCCAGGGTCTTAAGAACAAAGCCGATGCCACCGGTGATTGTGACAAAGATAG 60
QY 232 tctaacgtcctctagatcacacgttcttgaagtcaacctccggttcggacatcccgaggc 291
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CCCAACGTCGTAGATCTCATGTTCTTGAAGTCTCATCTGGAGCCGACATATAATTGAGAG 120
QY 292 agtctccactacgcccaactcgtcgcggtcgcggtttgcattataaagcgcacggtgc 351
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 CGTCAACAATTTACGTCGCCGAGAGAGAGAGGTGTCTCCATTTCTCAGTGTGTAACGGCGC 180
QY 352 ggtcactaacgtcacgatacacgcaacctcgcggtcgcggtg----- 393
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GGTGGCTAACCTCACCTCCGACAGCGGTGACGACTCATGGGAACAATGTGTGAATGA 240
QY 394 -----tgagggtgattaccctgcgtggttcggttcacattttgtctttgacccgg 444
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 AGCCGAGCTGGAGGATTTGACTTTACATGAAGGTTTGAGATTTCTTCCATCCTACTGG 300
QY 445 tactgcgttccacgcgcctgcaccacggagagaggtttgacggtgtatctagccgg 504
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 TACGGTGTCTCCGCCGCCGCCGCCGGATGCGGTGTTTTATCTATCTTTGTTGCTGG 360
QY 505 aggtcaagacaaagttgtaggaggaagtgtggttcgttaattgtctcggggaccggt 564
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 TGACAAGGTCCGATCATCGGAGGAGAGAGTGTGCTCGCTTGTGCTTCTGCTCCGCT 420
QY 565 agtgttgatggtcgtcttttgcacacgcagtttatgataggttaccgattgaagagga 624
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 GATACTGATGGTGTGATCTGTTCTCCAAACCAACTTTTGAAGGCTTCCGCTTTGAAGATGA 480
QY 625 agaa 628



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 19:01:52 ; Search time 1922.93 Seconds
(without alignments)
10599.682 Million cell updates/sec

Title: US-09-823-676-1
Perfect score: 974
Sequence: 1 cccccgacctgctctata.....cgaaatgcgaatattaggtt 974

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_hgt_hum.*

31: em_hgt_inv.*

32: em_hgt_other.*

33: em_hgtgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description

C	1	952.8	97.8	86710	8	ATF23E12	AL022604 Arabidops
	2	952.8	97.8	197859	8	ATCHRIV83	AL161587 Arabidops
C	3	212.4	21.8	1323	8	AF194974	AF194974 Arabidops
	4	212.4	21.8	103152	8	AC007369	AC007369 Arabidops
	5	208.6	21.4	88411	8	AC004667	AC004667 Arabidops
	6	205.2	21.1	122871	8	AC002387	AC002387 Arabidops
C	7	201.8	20.7	100806	8	AC015450	AC015450 Arabidops
	8	196.8	20.2	107600	8	ATF16J13	ATF16J13 Arabidops
C	9	196.8	20.2	190026	8	ATCHRIV33	AL049638 Arabidops
	10	191.6	19.7	197976	8	ATCHRIV39	AL161533 Arabidops
	11	191.6	19.7	206606	8	ATFCA1	AL161539 Arabidops
	12	190.4	19.5	1151	8	AF446359	297336 Arabidops
C	13	190.4	19.5	1151	8	AF446359	AF446359 Arabidops
	14	190.4	19.5	18835	8	ATFCA9	297344 Arabidops
C	15	186.4	19.1	198067	8	ATCHRIV47	AL161547 Arabidops
	16	186.4	19.1	89350	8	ATT12H17	AL021635 Arabidops
	17	183	18.8	129845	2	AP004165	AL161538 Arabidops
C	18	180.2	18.5	177418	2	OSJN00182	AP004165 Oryza sat
	19	179.4	18.4	107865	8	ATT4C21	AL662981 Oryza sat
C	20	178.8	18.4	103240	8	ATT22E16	AL162295 Arabidops
	21	177.2	18.2	95310	8	ATAC011437	AL132975 Arabidops
C	22	174	17.9	160363	2	AP003526	AC011437 Arabidops
	23	170.4	17.5	116205	8	AC006931	AC003526 Oryza sat
C	24	166	17.0	156834	2	AP004757	AC006931 Arabidops
	25	161.8	16.6	134712	2	AP003891	AP004757 Oryza sat
	26	161.8	16.6	135769	2	AP004587	AP003891 Oryza sat
	27	160.8	16.5	108763	2	AP004020	AP004587 Oryza sat
C	28	148.8	15.3	147666	2	AP004635	AP004020 Oryza sat
	29	148.8	15.3	152381	2	AP004698	AP004635 Oryza sat
C	30	127.6	13.1	12319	6	E64635	AP004698 Oryza sat
	31	127.6	13.1	71807	8	AB025613	E64635 Method for
	32	116.8	12.0	145847	2	AP004319	AB025613 Arabidops
	33	111.6	11.5	140466	8	AP003683	AP004319 Oryza sat
C	34	108.2	11.1	111686	8	F14L17	AP003683 Oryza sat
	35	101.2	10.4	279	8	AF184830	AC012188 Sequence
C	36	89.2	9.2	140327	2	AP003938	AF184830 Arabidops
	37	87.8	9.0	169656	2	AP004680	AP003938 Oryza sat
C	38	69.8	7.2	167764	2	AP004303	AP004680 Oryza sat
	39	50.8	5.2	125020	9	AF429315	AP004303 Oryza sat
C	40	45.2	4.6	590	8	HVU234403	AF429315 Homo sapi
	41	45.2	4.6	124050	8	AF474373	AF429315 Homo sapi
C	42	44.6	4.6	7218	6	I66494	AJ234403 Hordeum v
	43	43.8	4.5	125020	9	AF429315	AF474373 Hordeum v
	44	42.2	4.3	1137	8	ANAL12349	I66494 Sequence 14
	45	42	4.3	1694	8	ATAJ4119	AF429315 Homo sapi
							AJ132349 Antirrhin
							AJ224119 Arabidops

ALIGNMENTS

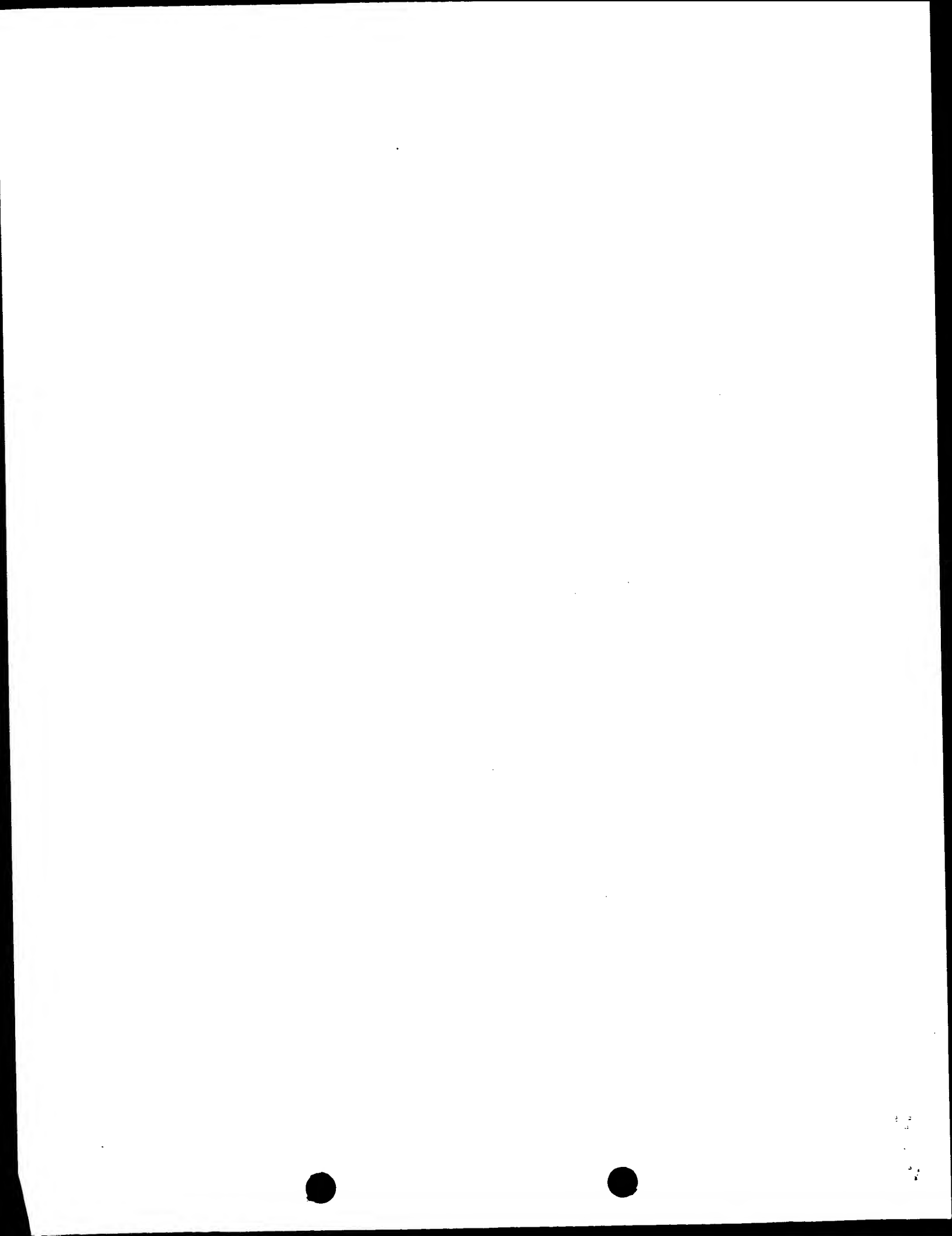
RESULT 1
ATF23E12/c
LOCUS
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F23E12 (BSSA
PROJECT)
VERSION AL022604.1
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana

REFERENCE
AUTHORS
JOURNAL
Unpublished
2 (bases 1 to 86710)
EU Arabidopsis sequencing, project.
Direct Submission
Submitted (01-APR-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail:

thale cress.
Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 86710)
Bevan, M., Hilbert, H., Braun, M., Holzer, E., Brandt, A.,
Duesterhoeft, A., Hohelsel, J., Mewes, H.W., Mayer, K.F.X. and
Schueller, C.

Unpublished
2 (bases 1 to 86710)
EU Arabidopsis sequencing, project.
Direct Submission
Submitted (01-APR-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail:



COMMENT

source

exon

exon

CD5

Lntron

axon

tron

Box

utron

xon

con-

ene

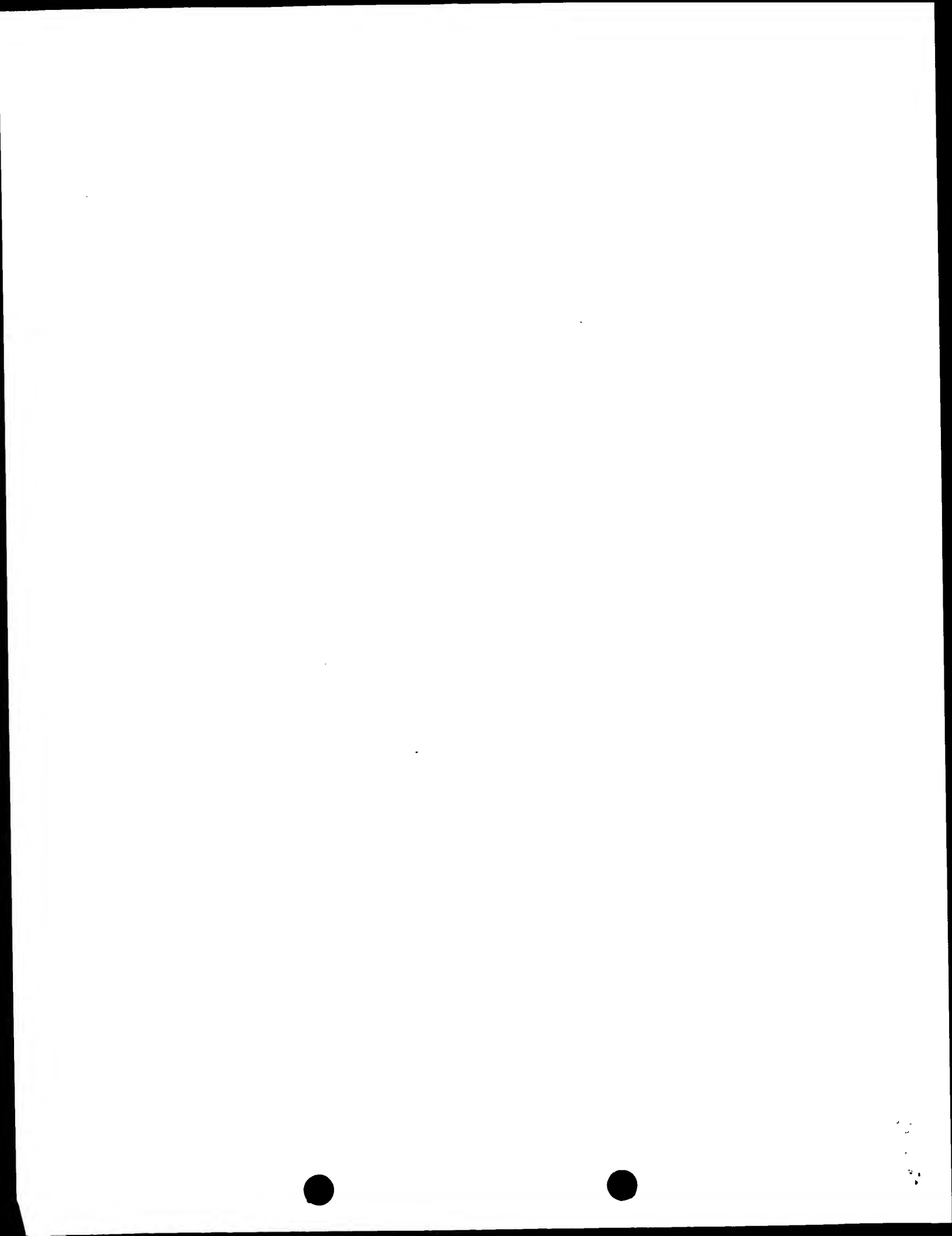
DS

intron

```

schuellemips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
Coordinator: John Innes, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@ebc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
Location/Qualifiers
1. .86710
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
173. .1890
/gene="F23E12.10"
173. .374
/gene="F23E12.10"
/number=1
join(<173. 374,456. 919,1089. .1610,1678. .1890)
/gene="F23E12.10"
/codon_start=1
/product="hypothetical protein (fragment)"
/protein_id="CAA18726.1"
/db_xref="GI:3080407"
/db_xref="SPTREMBL:O65486"
/translation="FNAVAGCFPAVESVLMPPSSSTSSLPNTTSMVILSAVATSV
VSEIGSGEPKFPDYDFRPGFGLVLLGALCVLSALSRCTSMTSYAVDSLNNK
DKIAPKAVPVNGSLVGYIALVPEVLYMGFQNDVILLEKRPVKGLSADLLQLVAPO
XAGIATWCRASGLVGYIAPSLFIGGAAGMAYGKFIGLALQNPDFNLSTLEVAPO
AYGLGMATLAGVQVPLTAVLLFELTQDYRIVPLIGAVGWSNITSGQSKRQET
RTETKRRKQSGEAVOSLTSSDDSSSTNNLCEVSSLCDDSLNQSEELPKSFVFEKA
NRPKDIYVMDICSRSGGCKKVPWIVPMDLAAQTIMNKHELHSAVVGSGSIDAPRI
HPGVLDRECIITLRR"
375. .455
/gene="F23E12.10"
/number=1
456. .919
/gene="F23E12.10"
/number=2
920. .1088
/gene="F23E12.10"
/number=2
1089. .1610
/gene="F23E12.10"
/number=3
1611. .1677
/gene="F23E12.10"
/number=3
1678. .1890
/gene="F23E12.10"
/number=4
3124. .3241
/gene="F23E12.20"
/number=1
3124. .4428
/gene="F23E12.20"
join(3124. .3241,3232. .3465,3595. .3696,3945. .4107,
4211. .4428)
/gene="F23E12.20"
/number=1
/note="similarity to dihydroflavonol-4-reductase, Gerbera
x sp., PIR2:395189"
/codon_start=1
/product="putative protein"
/protein_id="CAA18727.1"
/db_xref="GI:3080408"
/db_xref="SPTREMBL:O65487"
/translations="MDQAGKVCYTCASGFLASWLVKRLLLEGYEVIGTVRDPGNEKK
LAHLKLEGAKERLVKADLMEEGFDNAGMCQGFHTASPEILRPAIEGTLNVLV
SCRNPSSKRVLTSSSTVWYALSTLAEQAWKSEENGIDLVTLPSFLVGPSP
PDLCTASVDVGLLKGTEKTFQHWGQGVHIDDVARTHIVVFEHAAQGRYICSSN
ISLEELVSLFASRPSLTPKPKREDTDLN"
3242. .3232

```



```

Query Match          97.8%; Score 952.8; DB 8; Length 86710;
Best Local Similarity 99.7%; Pred. No. 2.le-258;
Matches 965; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

7  gacctgcctctacagagacctgaagattccagaaccccaacctgatacaaaataacatgga 66
   |||||
9099 GAACCTGCATCTACAGAGACCTTGAAGATTCAGAACCCACCCTGATC-AAATAATCAATGGA 9041

67  acttaacagatctgaagcagacgaacgaaggccgagaccactccaccggtggagccac 126
   |||||
9040 ACTTAAACAGATCTGAAGCAGACGAGCAAGCAAGGCCGAGACACTCCCACCGGTGGAGCCAC 8981

127  cagctcagccacagcctctggtctctccctccgagcgtcgatccacggtgcctcgagg 186
   |||||
8980 CAGCTCAGCCACAGCCTCTGGCTCTCCCTCCGGACGTCCTCCACGTGCTGCTCTGCAGG 8921

187  ttccaaaaaaccacaaacctccgacattataactagatagatgctctaagcctcttag 246
   |||||
8920 TTCCAAAAACAAACCAACCTCCGACGATTATAACTAGAGATAGTCTAACGTCCTTAG 8861

247  atcacagcttctgaagtcacctcgggtctcggacatccacgaggcagctccacctacgc 306
   |||||
8860 ATCACAGCTTCTTGAAGTGCACCTCCGGTTCGGACATATCCAGGCGAGTCTCCACCTACGC 8801

307  cactcgtcggcgtcgcggtttgcatataagcgcagcgggtgcggtcactaaagtcac 366
   |||||
8800 CACTCTGTCGGCGGTGCGCGCTTTCGATATACGGCGACGGGTGCGGTCACTAACGTCTAC 8741

367  gatacggcaacctcgcgctccgctggtggaagtggtattacctgcattggtcggtttga 426
   |||||
8740 GATACGCAACCTGCGCGCTCCGGCTGGTGGAGGTGTGATTACCTTCGATAGTTCGGTTGA 8681

427  catttgcctttgacgggtactgcgcttccaccgctcgcacaccggggagcaggagttt 486
   |||||
8680 CATTTTGTCTTTGACCGGTACTGCGCTTCCACCGCTTCACACCGGGGAGCAGGAGTTT 8621

487  gacggtgatctagccggagatcaaggacaaagtgtgagggaagtgtgctggttcgatt 546
   |||||
8620 GACGGTGTATCTAGCCGAGGTCAAGGACAAGTTGTAGAGGGNAATGTGCTGTTCTGTT 8561

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:19:08 ; Search time 103.68 Seconds
(without alignments)
450.508 Million cell updates/sec

Title: US-09-823-676-2

Perfect score: 1392

Sequence: 1 MELNRSEADEAKETPTGTG.....IYMGSGGGGGGATRPAP 270

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392	100.0	270	10	065489
2	706.5	50.8	311	10	0957C9
3	646.5	46.4	292	10	023620
4	645.5	46.4	302	10	09C9K7
5	637.5	45.8	315	10	09SR17
6	629.5	45.2	317	10	022130
7	625.5	44.9	339	10	09S270
8	616	44.3	285	10	082166
9	605	43.5	324	10	049662
10	581	41.7	265	10	09LZX7
11	572	41.1	310	10	09M2S3
12	554	39.8	257	10	09SJG4
13	499.5	35.9	276	10	09LTA2
14	417	30.0	206	10	09M9R4
15	407	29.2	265	10	0941Y9
16	289.5	20.8	404	10	091VB0

17	289	20.8	771	10	09XHY4
18	283	20.3	405	10	09AYM0
19	282	20.3	351	10	094F52
20	274	19.7	296	10	022812
21	274	19.7	351	10	023142
22	264.5	19.0	419	10	09FHM5
23	259	18.6	826	10	09SH33
24	258.5	18.6	300	10	092RR7
25	255.5	18.4	334	10	049658
26	248.5	17.9	386	10	09FIR1
27	246.5	17.7	347	10	004696
28	243	17.5	439	10	049350
29	243	17.5	439	10	094010
30	242.5	17.4	334	10	004695
31	239.5	17.2	404	10	09SB31
32	239	17.2	455	10	049694
33	227	16.3	365	10	09S273
34	225	16.2	348	10	080834
35	221	15.9	418	10	09SR16
36	218	15.7	345	10	081321
37	211.5	15.2	348	10	09M2D3
38	162.5	11.7	574	10	09SJO2
39	148	10.6	848	5	09V889
40	146	10.5	405	5	Q23057
41	143.5	10.3	818	6	Q9N1P0
42	139	10.0	1468	5	09GUB5
43	138.5	9.9	265	5	Q23347
44	137	9.8	694	16	053212
45	136.5	9.8	1610	16	092KQ8

ALIGNMENTS

RESULT 1

065489 ID 065489 PRELIMINARY; PRT: 270 AA.
AC 065489;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 26.8 KDA PROTEIN.
GN F23EL2.50 OR AT4G5390.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Hohelsel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.F.X.,
RA Schueller C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL022604; CAA18730.1; -
DR EMBL; AL161587; CAB80256.1; -
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 26774 MW; 4C21487C2719E114 CRC64;

Query Match 100.0%; Score 1392; DB 10; Length 270;
Best Local Similarity 100.0%; Pred. No. 4.7e-88;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

QY 1 MELNRSEADAKAETTPTGATSSATASGSSGRRPRGRPAGSKNKPPTTIITRSPNV 60
DB 1 MELNRSEADAKAETTPTGATSSATASGSSGRRPRGRPAGSKNKPPTTIITRSPNV 60
QY 61 LRSHVLEVTSGSDISEAVSTYATRRGGVCIISGTGAVTNVIRQAPAGGVITLHGR 120
DB 61 LRSHVLEVTSGSDISEAVSTYATRRGGVCIISGTGAVTNVIRQAPAGGVITLHGR 120
QY 121 FDILSLTGTALPPAPPAGGLTVYLAGGGQVVGNGVAGSLIAGSPVVLMAASFANAY 180
DB 121 FDILSLTGTALPPAPPAGGLTVYLAGGGQVVGNGVAGSLIAGSPVVLMAASFANAY 180
QY 181 DRUPIEETPTTPTGTVQOQPEASOSSEVTGSGAQACSNLQGGGGVAFYNLGMN 240
DB 181 DRUPIEETPTTPTGTVQOQPEASOSSEVTGSGAQACSNLQGGGGVAFYNLGMN 240
QY 241 MNPFQSGGDIYMGSGSGGGGGGATRPAPF 270
DB 241 MNPFQSGGDIYMGSGSGGGGGGATRPAPF 270

RESULT 2
ID Q9S7C9 PRELIMINARY; PRT; 311 AA.
AC Q9S7C9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ESCAROLA.
GN ESC OR F9H16.12.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Waigel D., Ahn J.H., Blazquez M.A., Borevitz J., Christensen S.K.,
RA Fankhauser C., Ferrandiz C., Kardailsky I., Neff M.M., Nguyen J.T.,
RA Sato S., Wang Z., Xia Y., Dixon R.A., Harrison M.J., Lamb C.J.,
RA Yanofsky M.F., Chory J.
RT "Activation Tagging in Arabidopsis."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA Lueros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
Walker M., Yu G., Ecker J., Theologis A., Davis R.W.
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF194974; AAF07197.1; -
DR EMBL; AC007369; AAD30602.1; -
DR InterPro; IPR000637; AT_hoek.
DR Pfam; PF02178; AT_hoek; 1.
DR SMART; SM00384; AT_hoek; 1.
SQ SEQUENCE 311 AA; 31842 MW; A80B445C9776EB7D CRC64;

Query Match 50.8%; Score 706.5; DB 10; Length 311;
Best Local Similarity 53.0%; Pred. No. 5e-41;
Matches 151; Conservative 25; Mismatches 56; Indels 53; Gaps 6;

QY 7 EADAKAETTPTGATSSATASGSSGRRPRGRPAGSKNKPPTTIITRSPNLSHVL 66
DB 59 ESDSHNDHQQQRDPDPTSSAPGKPRGRPPGSKNKAAPPIIVTRDSPNLSHVL 118
QY 67 EVTSGSDISEAVSTYATRRGGVCIISGTGAVTNVIRQAPAP-----AGGGVITLHG 119
DB 119 EVSPGADIVESVYARRRGVSVLGGNGTVSNVIRQVPTPGNGGGVGGGGVITLHG 178
QY 120 RFDILSLTGTALPPAPPAGGLTVYLAGGGQVVGNGVAGSLIAGSPVVLMAASFANAY 179

```

```

DB 179 RFEILSLTGTALPPAPPAGGLSIFLAGGGQVVGSSVAPLIASAPVILMAASNAV 238
QY 180 YDRUPIEETPTTPTGTVQOQPEASOSSEVTGSGAQACSNLQGGNG 228
DB 239 FERUPIEETPTTPTGTVQOQPEASOSSEVTGSGAQACSNLQGGNG 287
QY 229 GGGVAFYNLGMNMFQSGGDIYMGSGS-----GGGGGGGATRPAPF 270
DB 288 GG-----YGFSGDPHLLGWGAGTSPRPFP 311

RESULT 3
ID O23620 PRELIMINARY; PRT; 292 AA.
AC O23620;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 29.7 KDA PROTEIN.
GN AT4G17800.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98121113; PubMed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dikse W., van Staveren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Wedler E., Wambutt R., Weitzenecker T., Pohl T.M., Terry N.,
RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Auborg S., Gy I., Kreis M., Kavanagh T., Hempel S., Kotter P.,
RA Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Silvey M., James K., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Milloni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Rechman S., Ansoorge W., Cooke R., Berger C.,
RA Delsen M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chalwatzis N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana."
RL Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97344; CAB10560.1; -
DR EMBL; AL161547; CAB78783.1; -
KW Hypothetical protein.
SQ SEQUENCE 292 AA; 29724 MW; 2A948700FA543C4F CRC64;

Query Match 46.4%; Score 646.5; DB 10; Length 292;
Best Local Similarity 52.7%; Pred. No. 5.9e-37;
Matches 136; Conservative 32; Mismatches 47; Indels 43; Gaps 8;

QY 19 GQATSSATASGSS-----GRPRGRPAGSKNKPPTTIITRSPNLSHVLVTSQSD 73
DB 62 GGGSGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 121
QY 74 ISEAVSTYATRRGGVCIISGTGAVTNVIRQAPAGGVITLHGRDILSLTGTALPP 133
DB 122 VEDCVATARRQRQICVLGSGVTVNVIIRQPS--AGAVVTLOGTEILSLSGSFLPP 179
QY 134 PAPPAGGAGTIVYLAGGGQVVGNGVAGSLIAGSPVVLMAASFANAYDRLPPIEETPPP 193
DB 180 PAPPAGTSLTIFLAGGGQVVGNGVAGSLIAGSPVVLMAASFANAYERLPLEEDE---- 235
QY 194 RTTGTVQOQPEASOSSEVTGSGAQACSNLQGGN-----GGGVAFYNLGMNMF 244

```


PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 18-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139763.
 PR 22-JUN-1999; 99US-0139817.
 PR 23-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 06-JUL-1999; 99US-0142055.
 PR 08-JUL-1999; 99US-0142390.
 PR 09-JUL-1999; 99US-0142803.
 PR 12-JUL-1999; 99US-0142820.
 PR 13-JUL-1999; 99US-0142977.
 PR 14-JUL-1999; 99US-0143542.
 PR 15-JUL-1999; 99US-0143624.
 PR 16-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 20-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0148684.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 23-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 25-AUG-1999; 99US-0149930.
 PR 26-AUG-1999; 99US-0150566.
 PR 27-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151080.
 PR 31-AUG-1999; 99US-0151303.
 PR 01-SEP-1999; 99US-0151303.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.

```

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.

Query Match 45.8%; Score 637.5; DB 21; Length 280;
Best Local Similarity 50.0%; Pred. No. 9.4e-44;
Matches 132; Conservative 37; Mismatches 58; Indels 37; Gaps 5;

QY 17 PTGGATSSATASGSSRRPRGPKPKPTIITRDSPNVLRSHVLEVTSGDISSE 76
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |
Db 35 pregaveapt-----rrprgpragsknkpkpifvtrdspnalkshvmeiasgtdiv 87
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |
QY 77 AVSTYATRRGCGVCIISGTGAVTNVTIRQP-----AAPAGGGVITLHGRFDILSLGTA 130
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |
Db 88 tlatfarrqrgicilsgngtvantlrqstaavaapgaavialqgffellslcgf 147
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |
QY 131 LPPAPPAGAGLTVYLAGGGQGVYGVGNVAGSLIAGSPVVLMAASFANAVYDRLPFIEEET 190
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |
Db 148 lpgpappgsgtltiylagggqgvvggsvvplmaagpvmliatfsnatyerlpleeaa 207
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |
QY 191 PP----PRTTGVOQQPPEASOSSEVTGSGAOACESNLOGNGGGGVAFYNLGMNMNFQF 246
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |
Db 208 aerggggsggvvpq-----lgggsgplssggaggdngdglpvympgnlvs--- 255
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |
QY 247 SGGDIYGMSSGGGGGGGATRPAPF 270
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |
Db 256 -----ngsggggqmsgdeay 271
   ||| : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||| : |

RESULT 6
AAG10649
ID AAG10649 standard; Protein; 282 AA.
XX AC AAG10649;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 9055.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125786.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
```


PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 20-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144684.	PR 18-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161921.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161992.
PR 04-AUG-1999;	99US-0147204.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147202.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		

Query Match 46.4%; Score 646.5; DB 21; Length 292;
 Best Local Similarity 52.7%; Pred. No. 1.9e-44;
 Matches 136; Conservative 32; Mismatches 47; Indels 43; Gaps 8;

Db 241 mnfnfsggdiymsgsgggggtatpaf 270
RESULT 3
AAG29345
ID AAG29345 standard; Protein; 254 AA.
XX
AC AAG29345;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 34900.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

PA (KEDD/) KEDDIE J.
 PA (RATC/) RATCLIFFE O.
 PA (HEAR/) HEARD J.
 PA (SAMA/) SAMAHA R.
 PA (YUGG/) YU G.
 PA (JIAN/) JIANG C.
 XX
 PI Riechmann JL, Reuber L, Keddie J, Ratcliffe O, Heard J, Samaha R;
 PI Yu G, JIANG C;
 XX
 XX WPT: 2001-335996/35.
 DR N-PSDB: AAD05839.
 XX
 PT Nucleic acids encoding plant transcription factor polypeptides, useful
 PT for altering the developmental and structural characteristics of
 PT plants, e.g. corn, potato and cotton plants -
 XX
 XX Claim 4; Page 103-104; 140pp; English.
 XX
 CC The present sequence is Arabidopsis thaliana transcription factor G1073.
 CC The transcription factors are used to modify traits associated
 CC with structural or developmental characteristics of plants, e.g. soybean,
 CC wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa,
 CC sugar cane, turf, banana, blackberry, blueberry, strawberry, raspberry,
 CC cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes,
 CC honey dew, lettuce, mango, melon, papaya, peas, watermelon, pineapple,
 CC spinach, squash, sweet corn, tobacco, tomato, peppers, onion, roseaceous
 CC fruits and/or vegetable brassicas when their expression levels are
 CC altered. The manipulation of transcription factor levels in plants
 CC offers great potential in agricultural biotechnology for modifying
 CC plant's traits. The transcription factors are also used in gene therapy.
 XX
 SQ Sequence 270 AA;

Query Match 100.0%; Score 1392; DB 22; Length 270;
 Best Local Similarity 100.0%; Pred. No. 8.6e-105;
 Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MELNRSEADEAKAETPTGGATSSATASGSSGRRPRGRPAGSKNPKPPTIITRDSNV 60
 DB 1 melnrseadeakaetptggatssatassgrrprgrpagsknkpptiitrdspnv 60
 QY 61 LRSNVLVETSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
 DB 61 lrshvlevtsgsdiseavstyatrgrcgvcilsgtgavtnvtirqpapagggvithgr 120
 QY 121 FDILSLTGTALPPAPPAGGLTVYLAGGQGVVGGNVAGSLIASGPPVLMASAFANAVY 180
 DB 121 fdilsltgtalppappaggltvylaggqgvvvggnvagsliasgppvvlmaasfanavy 180
 QY 181 DRLPIEBEETPPPTTGVQOQPEASQSSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
 DB 181 drlpieeetppprtgvqqqqpeasqsssevtgsgaqacesnlqggnggggvafynlgnm 240
 QY 241 MNFQFSGGDIYGMSSGGGGGATRPAP 270
 DB 241 mnfnfsggdiygmsgggggatrpa 270

RESULT 2
 AAB67838
 ID AAB67838 standard; Protein; 270 AA.
 XX
 AC AAB67838;
 XX
 DT 29-JUN-2001 (first entry)
 XX
 DE Amino acid sequence of a plant transcription factor G1073.
 XX
 KW Transcription factor; flowering time; transgenic plant; vernalisation;
 KW plant development; plant physiology; flowering.
 XX

OS Arabidopsis thaliana.
 XX
 PN WO200126459-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 11-OCT-2000; 2000WO-US28141.
 XX
 PR 12-OCT-1999; 99US-0159464.
 PR 08-NOV-1999; 99US-0164132.
 PR 17-NOV-1999; 99US-0166228.
 PR 17-APR-2000; 2000US-0197899.
 PR 22-AUG-2000; 2000US-0227439.
 XX
 PA (MEND-) MENDEL BIOTECHNOLOGY INC.
 PA (RATC/) RATCLIFFE O.
 PA (HEAR/) HEARD J.
 PA (SAMA/) SAMAHA R.
 PA (CREE/) CREELMAN R.
 PA (KEDD/) KEDDIE J.
 PA (JIAN/) JIANG C.
 PA (REUB/) REUBER L.
 PA (RIEC/) RIECHMANN J L.
 XX
 PI Ratcliffe O, Heard J, Samaha R, Creelman R, Keddie J, JIANG C;
 PI Reuber L, Riechmann JL;
 XX
 XX WPI: 2001-266398/27.
 DR N-PSDB: AAF80408.
 XX
 PT New transgenic plant comprises a recombinant polynucleotide encoding a
 PT plant transcription factor polypeptide and has a modified flowering a
 PT time or vernalization requirement -
 XX
 XX Example 7; Page 94-95; 108pp; English.
 XX
 CC The present sequence represents a plant transcription factor protein
 CC which modifies the flowering time of a plant. The polynucleotide
 CC sequence is used to produce transgenic plants which have a modified
 CC flowering time or a modified vernalisation requirement. The
 CC polynucleotides and polypeptides are useful for modifying plant
 CC development, physiology or biochemistry such that the modified plants
 CC have a trait advantage over wild type plants. In particular they are
 CC useful for accelerating, delaying or preventing flowering. The
 CC polynucleotides are also useful as nucleic acid probes and primers.
 CC They may be used to identify proteins that can modify the activity of
 CC the transcription factor.
 XX
 SQ Sequence 270 AA;

Query Match 100.0%; Score 1392; DB 22; Length 270;
 Best Local Similarity 100.0%; Pred. No. 8.6e-105;
 Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MELNRSEADEAKAETPTGGATSSATASGSSGRRPRGRPAGSKNPKPPTIITRDSNV 60
 DB 1 melnrseadeakaetptggatssatassgrrprgrpagsknkpptiitrdspnv 60
 QY 61 LRSNVLVETSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
 DB 61 lrshvlevtsgsdiseavstyatrgrcgvcilsgtgavtnvtirqpapagggvithgr 120
 QY 121 FDILSLTGTALPPAPPAGGLTVYLAGGQGVVGGNVAGSLIASGPPVLMASAFANAVY 180
 DB 121 fdilsltgtalppappaggltvylaggqgvvvggnvagsliasgppvvlmaasfanavy 180
 QY 181 DRLPIEBEETPPPTTGVQOQPEASQSSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
 DB 181 drlpieeetppprtgvqqqqpeasqsssevtgsgaqacesnlqggnggggvafynlgnm 240
 QY 241 MNFQFSGGDIYGMSSGGGGGATRPAP 270
 DB 241 mnfnfsggdiygmsgggggatrpa 270

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 21:20:07 ; Search time 69.12 seconds
(without alignments)
433.882 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNSEADEAKAETPTGG.....IYMGSGGGGGGATRPAP 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1392	100.0	270	AAE01961	Arabidopsis thalia
2	1392	100.0	270	AAE01961	Amino acid sequenc
3	646.5	46.4	254	AAE01961	Arabidopsis thalia
4	646.5	46.4	254	AAE01961	Arabidopsis thalia
5	637.5	45.8	280	AAE01961	Arabidopsis thalia
6	637.5	45.8	280	AAE01961	Arabidopsis thalia
7	637.5	45.8	280	AAE01961	Arabidopsis thalia
8	625.5	44.9	339	AAE01961	A. thaliana transc
9	572	41.1	272	AAE01961	Arabidopsis thalia
10	572	41.1	272	AAE01961	Arabidopsis thalia
11	572	41.1	310	AAE01961	Arabidopsis thalia

12	572	41.1	310	21	AAE01961	Arabidopsis thalia
13	572	41.1	315	21	AAE01961	Arabidopsis thalia
14	572	41.1	316	21	AAE01961	Arabidopsis thalia
15	570	40.9	239	21	AAE01961	Arabidopsis thalia
16	570	40.9	239	21	AAE01961	Arabidopsis thalia
17	570	40.9	267	21	AAE01961	Arabidopsis thalia
18	570	40.9	267	21	AAE01961	Arabidopsis thalia
19	570	40.9	281	21	AAE01961	Arabidopsis thalia
20	570	40.9	281	21	AAE01961	Arabidopsis thalia
21	274	19.7	336	21	AAE01961	Arabidopsis thalia
22	274	19.7	343	21	AAE01961	Arabidopsis thalia
23	274	19.7	351	21	AAE01961	Arabidopsis thalia
24	255.5	18.4	304	21	AAE01961	Arabidopsis thalia
25	255.5	18.4	304	21	AAE01961	Arabidopsis thalia
26	255.5	18.4	334	21	AAE01961	Arabidopsis thalia
27	255.5	18.4	334	21	AAE01961	Arabidopsis thalia
28	255.5	18.4	334	22	AAE01961	Arabidopsis thalia
29	249	17.9	272	21	AAE01961	Arabidopsis thalia
30	249	17.9	272	21	AAE01961	Arabidopsis thalia
31	248.5	17.9	386	22	AAE01961	A. thaliana transc
32	243	17.5	386	22	AAE01961	A. thaliana transc
33	234.5	16.8	356	22	AAE01961	Arabidopsis thalia
34	233.5	16.8	277	21	AAE01961	Arabidopsis thalia
35	233.5	16.8	352	21	AAE01961	Arabidopsis thalia
36	233.5	16.8	352	21	AAE01961	Arabidopsis thalia
37	227	16.3	290	21	AAE01961	Arabidopsis thalia
38	227	16.3	365	21	AAE01961	Arabidopsis thalia
39	227	16.3	369	21	AAE01961	Arabidopsis thalia
40	225	16.2	264	21	AAE01961	Arabidopsis thalia
41	225	16.2	264	21	AAE01961	Arabidopsis thalia
42	225	16.2	342	21	AAE01961	Arabidopsis thalia
43	225	16.2	342	21	AAE01961	Arabidopsis thalia
44	225	16.2	348	21	AAE01961	Arabidopsis thalia
45	225	16.2	348	21	AAE01961	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAE01961
ID AAE01961 standard; Protein: 270 AA.
XX AAE01961;
XX AAE01961;
DT 31-JUL-2001 (first entry)
XX Arabidopsis thaliana transcription factor G1073.
DE Transcription factor; trait modification; structural characteristic;
KW developmental characteristic; gene therapy; agricultural biotechnology.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT Domain 33..42 /note= "Conserved domain"
FT Domain 78..175 /note= "Conserved domain"
FT Domain 78..175 /note= "Conserved domain"
XX WO200136444-A1.
XX 25-MAY-2001.
XX 14-NOV-2000; 2000WO-US31325.
XX 17-NOV-1999; 99US-0166228.
PR 17-APR-2000; 2000US-0197899.
PR 22-AUG-2000; 2000US-0227439.
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (RIEC/) RIECHMANN J L.
PA (REUB/) REUBER L.

FT CARBOHYD 81 81 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 83 83 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 87 87 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 91 91 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 93 93 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 94 94 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 96 96 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 98 98 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 101 101 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 103 103 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 104 104 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 106 106 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 107 107 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 108 108 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 110 110 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 111 111 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 114 114 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 123 123 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 124 124 O-LINKED (GALNAC. . .) (MUCIN TYPE).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 917 917 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 985 985 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1002 1002 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1068 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1150 AA; 109615 MW; 3CB68B5D29DD7F5A CRC64;

Query Match 9.2%; Score 128.5; DB 1; Length 1150;
 Best Local Similarity 26.6%; Pred. No. 0.47;
 Matches 79; Conservative 30; Mismatches 107; Indels 81; Gaps 17;

QY 18 TGGATSSATSSGSSRRPRGPKPKPIITRDSPNVRSHVLEVTGSDISEA 77
 DB 13 TGTGSGAGSTGSSG--STGATGASIGQPETSRISVAGS-----SGAPAVSSGA--SQA 63
 QY 78 VSTYATRRGCGVCIIISGT---AVNTVTIROPAP--AGGCVI-TLHGRFDILSLTGTA 131
 DB 64 AGT-----SGAGPCTTASSVGVETARPSPVAGSGTTGTGVS-----ASGSG 105
 QY 132 PPPAPPAGGLTVYLAGGQGVGNVAGSLIAGPVLMAAFANAVYDRLP----- 184
 DB 106 SSSGSPGATGASI-----GQPETSRISVAGS--SGAPAVSSGASGAGTSCAGPCTTASSV 159
 QY 185 -IEEETP-----PPTTGVOQQQPEASOSSEVTGSGAQACE 220
 DB 160 GVTEARPSVAGSTTTGVSAGSGTSSGSGPGATGATGQTPETSRISVAGSGGAPAVS 219
 QY 221 SNLQ---GGNGGG-GVAFYNLGMNMN---NFQPSG--GDIYMSG--GSGGGGGGAT 266
 DB 220 SGASQAAGTSCAGPCTTASSVGVETARPSPVAGSGTTGTGVSAGSGTSSGSGPGAT 276

RESULT 3
 FSH_DROME STANDARD; PRT; 2038 AA.
 AC P13709; P13710;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Female sterile homeotic protein (Fragile-chorion membrane protein).
 GN FS(1)H OR FSH.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89276730; PubMed=2567251;
 RA Haynes S.R., Mozer B.A., Bhatia-Dey N., Dawid I.B.;
 RT "The Drosophila fish locus, a maternal effect homeotic gene, encodes
 apparent membrane proteins.";

RL Dev. Biol. 134:246-257 (1989).
 CC -1- FUNCTION: REQUIRED MATERIALLY FOR PROPER EXPRESSION OF OTHER
 CC HOMOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
 CC -1- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
 CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M23221; AAA28540.1; .
 DR EMBL; M23222; AAA28541.1; ALT_TERM.
 DR EMBL; M15762; AAA70424.1; .
 DR EMBL; M15763; AAA70423.1; .
 DR EMBL; M15764; AAA70422.1; .
 DR PIR; A43742; A43742.
 DR HSP; P04002; LWFA.
 DR FlyBase; Fggn0004556; fs(1)h.
 DR InterPro; IPR001487; Bromodomain.
 DR Pfam; PF00439; bromodomain; 2.
 DR PRINTS; P00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 2.
 DR PROSITE; PS00633; BROMODOMAIN_1; 2.
 DR PROSITE; PS50014; BROMODOMAIN_2; 2.
 KW Developmental protein; Bromodomain; Transmembrane; Repeat.
 FT DOMAIN 51 123 BROMODOMAIN 1.
 FT DOMAIN 495 567 BROMODOMAIN 2.
 FT DOMAIN 945 1106 ET DOMAIN.
 FT TRANSMEM 330 350 POTENTIAL.
 FT TRANSMEM 451 471 POTENTIAL.
 FT TRANSMEM 750 770 POTENTIAL.
 FT TRANSMEM 790 810 POTENTIAL.
 FT TRANSMEM 816 830 POTENTIAL.
 FT TRANSMEM 874 894 POTENTIAL.
 FT TRANSMEM 1731 1751 POTENTIAL.
 FT TRANSMEM 1939 1959 POTENTIAL.
 FT VARIANT 909 909 G -> A; RKPY.
 FT VARIANT 1022 1022 H -> RKPY.
 SQ SEQUENCE 2038 AA; 205332 MW; 849E0706D50A0098 CRC64;

Query Match 9.0%; Score 125.5; DB 1; Length 2038;
 Best Local Similarity 23.1%; Pred. No. 1.3; Mismatches 31; Indels 101; Gaps 13;
 Matches 73; Conservative 31; Mismatches 31; Indels 101; Gaps 13;
 QY 7 EADEAKAETPTTGGATSSA-----TAGSSSSRRPRGRPGAS-----KNKPKPTIITR 55
 DB 1633 KANEQNLKNASSWSLANSPOSHTSSSSSS---KAKPAMDSFQFRNKA-----ER 1684
 QY 56 DSPNVLRSHVLEVTSGSDISEAVSYATRRGCGVCIISGTCAVTNVTIRQPAAPA--GG 112
 DB 1685 DRKLLEAAEKEKKEAEEQKQKHKK-----SSSSSLTSAVAQAATAA 1737
 QY 113 GVTILHGRFDILSLTGTAIPPPAPPAGGLTVYLAGGQGVV----- 154
 DB 1738 AAVTL-GAAAAAALASSANSNPGSGSGG-----AGSTSQQAATGDRDRDRDRERERS 1791
 QY 155 -----GNNVAGSLIASGPVVLMAAFANAVYDRLPIEETTPPTTGTGVOQQQPEA 205
 DB 1792 GSGGGQSGNGNNSNSNSNGSPGSAGSG-----SGGGGSGGA 1830
 QY 206 SOSSEVTGSGAQACSNLQGGNGCGGVAFYNLGMNMN-----NFQPSGSDIYGM 254
 DB 1831 SAGGPNSSGGGTA-NSNSGGGGGGGGPALLNAGNSNSGSGGAASNSNSVGVGIVGS 1889
 QY 255 SG-----GSGGGGGG 264
 DB 1890 GPGGNSGSGSGGGGG 1905

```

RESULT 4
MEF2_DROME
ID MEF2_DROME STANDARD; PRT; 515 AA.
AC P40791;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Myocyte-specific enhancer factor 2.
GN MEF2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94261646; PubMed=8202544;
RA Lilly B., Galewsky S., Firulli A.B., Schulz R.A., Olson E.N.;
RT "D-MEF2: a MADS box transcription factor expressed in differentiating
RT mesoderm and muscle cell lineages during Drosophila embryogenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:5662-5666(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94329547; PubMed=8052612;
RA Nguyen H.T., Bodmer R., Ammayr S.M., McDermott J.C., Spoerel N.A.;
RT "D-mef2: a Drosophila mesoderm-specific MADS box-containing gene with
RT a biphasic expression profile during embryogenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7520-7524(1994).
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT COULD BE A KEY PLAYER IN EARLY
CC MESODERM DIFFERENTIATION AND MAY BE REQUIRED FOR SUBSEQUENT CELL
CC FATE SPECIFICATIONS WITHIN THE SOMATIC AND VISCERAL/HEART
CC MESODERMAL LAYERS.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: MESODERM.
CC -!- DEVELOPMENTAL STAGE: FIRST DETECTABLE IN THE PRESUMPTIVE MESODERM
CC AT LATE CELLULAR BLASTODERM STAGE. EXPRESSED IN ALL PRESUMPTIVE
CC MESODERM PRIOR TO THE SPLITTING PROCESS THAT GENERATES THE SOMATIC
CC AND VISCERAL/ HEART MESODERM AFTER THE SUBDIVISION, IT IS FOUND
CC IN BOTH THE SOMATIC AND THE VISCERAL/HEART MESODERM.
CC -!- INDUCTION: TWI ACTIVITY IS REQUIRED FOR MEF2 EXPRESSION. SNA
CC ACTIVITY IS NEEDED FOR MAINTAINING MEF2 EXPRESSION.
CC -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC FACTORS. MEF2 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U03292; AAA19957.1; -
CC EMBL; U07422; AAA20463.1; -
CC HSP; P11831; ISRS.
CC TRANSFAC; T01772; -
CC FlyBase; Fgn0011656; Mef2.
CC InterPro; IPR002100; MADS-box.
CC Pfam; PF00319; SRF-TF; 1.
CC PRINTS; PR00404; MADSDOMAIN.
CC SMART; SM00432; MADS; 1.
CC PROSITE; PS00350; MADS_BOX_1; 1.
CC PROSITE; PS00666; MADS_BOX_2; 1.
CC Transcription regulation; DNA-binding; Nuclear protein;
CC Developmental protein; Differentiation; Activator.
CC
CC DOMAIN 3 57 MADS.
CC FT DNA_BIND 58 86 MEF2-TYPE (POTENTIAL).
CC FT DOMAIN 380 386 POLY-SER.
CC FT DOMAIN 390 396 POLY-GLY.
CC FT DOMAIN 421 428 POLY-GLY.

```

```

FT DOMAIN 484 495 POLY-GLN.
FT CONFLICT 338 338 S -> R (IN REF. 2).
FT CONFLICT 364 377 PAVISIRIAMVPRAG -> ASGHQONSNGSTGS (IN
FT REF. 2).
SQ SEQUENCE 515 AA; 54289 MW; 832C1FB4A831E28E CRC64;

Query Match 8.9%; Score 123.5; DB 1; Length 515;
Best Local Similarity 23.1%; Pred. No. 0.43;
Matches 66; Conservative 31; Mismatches 114; Indels 75; Gaps 12;

QY 16 TPFGGATSSATASGS-----SSGRRPRGRPAGSKNKPPTIITRDSFNVLRSHVLEVTSG 71
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Dy 184 TDSGGMSLIITYPSGMLSENGPHSHSPLVGSFSP-----SPGI--AHLSIKQ 234
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 72 SDISEAVSTYATRRGCGVCIIISGTGAVTVNTIROP-----AAPAGGVITLHGRF 121
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Dy 235 SPGSQ-----NGRASNLRVVIPPPIPPMMSAPDDVGADQRQSQ 276
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 122 DILSLTCTALPPPPAGGAGLTIVLAGGQGVVGNVAGSLIAGFPVLMMAASFANAVYD 181
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Dy 277 TSLNTPVVTLTQTP-----ALTSYFCAQ-----DFSSGVNMSADIMSLNTHOGLVP 326
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 182 RLPIE-----EEETPPPTTCG-----OQOQPEASQSEVT-----GSGAQACES 221
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Dy 327 HSSLHLAVNSNTPPPATSPVSIKVAEPQSPRPDLSPAVISRIAMVPRAGGSSSSSTSS 386
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 222 NLOGGGGGG-VAFYNLGMNMFQFSGDIYCMGSGGGGGGGAT 266
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Dy 387 NASGAGGGGGAASAAVITHLNNVSLAG---GPSGGGGGGGGGS 429
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 5
ANDR_EULFC STANDARD; PRT; 884 AA.
ID ANDR_EULFC
AC O97776;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Androgen receptor (Dihydrotestosterone receptor).
GN AR OR NR3C4.
OS Eulemur fulvus collaris (Collared brown lemur).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Lemuridae; Eulemur.
OX NCBI_TaxID=47178;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98404153; PubMed=9732460;
RA Choong C.S., Kempainen J.A., Wilson E.M.;
RT "Evolution of the primate androgen receptor: a structural basis for
RT disease.";
RL J. Mol. Evol. 47:334-342(1998).
CC -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR3 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U94178; AAC73049.1; -
CC HSP; P06536; IGDC.
CC InterPro; IPR001103; Androgen_recep.
CC InterPro; IPR000536; Hormone_rec_lig.

```

DR InterPro: IPR001628; zf-C4.
 DR Pfam: PF02166; Androgen_recep; 1.
 DR Pfam: PF00104; hormone_rec; 1.
 DR Pfam: PF00105; zf-C4; 1.
 DR PRINTS: PR00047; STROIDFINGER.
 DR SMART: SM00430; HOLI; 1.
 DR SMART: SM00399; ZnF_C4; 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 KW Receptor; transcription regulation; DNA-binding; Nuclear protein;
 KW Zinc-finger; Steroid-binding.
 FT DOMAIN 1 522 MODULATING (BY SIMILARITY).
 FT DNA_BIND 524 589 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 524 544 C4-TYPE.
 FT ZN_FING 560 584 C4-TYPE.
 FT DOMAIN 655 884 LIGAND-BINDING.
 FT DOMAIN 55 58 POLY-GLN.
 FT DOMAIN 64 70 POLY-GLN.
 FT DOMAIN 116 120 POLY-ALA.
 FT DOMAIN 174 178 POLY-GLN.
 FT DOMAIN 353 362 POLY-PRO.
 FT DOMAIN 379 383 POLY-ALA.
 FT DOMAIN 408 411 POLY-ALA.
 FT DOMAIN 430 435 POLY-GLY.
 SQ SEQUENCE 884 AA; 95610 MW; 18F570E352F4D2BD CRC64;

Query Match 8.8%; Score 123.5; DB 1; Length 884;
 Best Local Similarity 25.1%; Pred. No. 0.74;
 Matches 76; Conservative 28; Mismatches 136; Indels 63; Gaps 12;
 QY 1 MELNRSEADEAKAETPTGGATSSATASGSSG--RRPRGRPGSKNKPPTIITRDS 58
 Db 170 MQLQQQQQEAUSE-----GSSGRAREAGAPTSSKDSYLCGGTSTSDSA 215
 QY 59 NVLRSHV-LEVTSGSDISEAVSTYATRRG-CGVCIISGTGAVNTVIRQPAAPAGGGVIT 116
 Db 216 KELCAKAVSVSMGLGVETLEHLSPEQLRGDCMYPVAVRPTCAPLAECCKGSL- 274
 QY 117 LHGRFDILSLGTALPPPPAGGLTVYLAGGQGVVGVGNVAG-----SLIAS 165
 Db 275 -----DSADKGTETPEAYTFPGSYTQGLGESLGSSEAGSGTLELPSTLSLYKS 329
 QY 166 GPVYLMAASFANAVYDRLPIDEEETPPRPT-----TGVOQQQPEASQSEVTSQAQCE 220
 Db 330 G-ALAEAAQYQSRDYNYNPLALAGPPPPPPHARIKLENP-LDYGSSWAAAAAQCRRF 387
 QY 221 SNLGGNGGGGV-----AFYNLGMNMNPFQSGGDIYGMSSGGGG--GGGA 265
 Db 388 GDLSLHGGGATPGSGSPSAAASSWHTL-----FTAEGLYPCGGGGGTSEAGA 441
 QY 266 TRP 268
 Db 442 VTP 444

RESULT 6
 ANDR_MACFA STANDARD; PRT; 895 AA.
 AC 097952;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Androgen receptor (Dihydrotestosterone receptor).
 GN AR OR NR3C4.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98404153; PubMed=9732460;
 RA Choong C.S., Kempainen J.A., Wilson E.M.;

Evolution of the primate androgen receptor: a structural basis for disease;
 J. Mol. Evol. 47:334-342(1998).
 CC -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN, A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NR3 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U94179; AAC73050.1; -.
 DR HSPB: P06536; LGDC.
 DR InterPro: IPR001103; Androgen_recep.
 DR InterPro: IPR000536; Hormone_rec_lig.
 DR InterPro: IPR001628; zf-C4.
 DR Pfam: PF02166; Androgen_recep; 1.
 DR Pfam: PF00104; hormone_rec; 1.
 DR Pfam: PF00105; zf-C4; 1.
 DR PRINTS: PR00047; STROIDFINGER.
 DR SMART: SM00430; HOLI; 1.
 DR SMART: SM00399; ZnF_C4; 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 KW Zinc-finger; Steroid-binding.
 FT DOMAIN 1 533 MODULATING (BY SIMILARITY).
 FT DNA_BIND 535 600 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 535 555 C4-TYPE.
 FT ZN_FING 571 595 C4-TYPE.
 FT DOMAIN 666 895 LIGAND-BINDING.
 FT DOMAIN 55 62 POLY-GLN.
 FT DOMAIN 68 74 POLY-GLN.
 FT DOMAIN 178 182 POLY-GLN.
 FT DOMAIN 357 366 POLY-PRO.
 FT DOMAIN 381 387 POLY-ALA.
 FT DOMAIN 434 448 POLY-GLY.
 SQ SEQUENCE 895 AA; 96494 MW; A3EB17916F43A097 CRC64;

Query Match 8.8%; Score 122.5; DB 1; Length 895;
 Best Local Similarity 23.4%; Pred. No. 0.86;
 Matches 70; Conservative 35; Mismatches 129; Indels 65; Gaps 11;
 QY 1 MELNRSEADEAKAETPTGGATSSATASGSSG--RRPRGRPGSKNKPPTIITRDS 58
 Db 174 MQLQQQQQEAUSE-----GSSGRAREAGAPTSSKDSYLCGGTSTSDSA 219
 QY 59 NVLRSHV-LEVTSGSDISEAVSTYATRRG-CGVCIISGTGAVNTVIRQPAAPAGGGVIT 116
 Db 220 KELCAKAVSVSMGLGVETLEHLSPEQLRGDCMYPVAVRPTCAPLAECCKGSL- 278
 QY 117 LHGRFDILSLGTALPPPPAGGLTVYLAGGQGVVGVGNVAGSLIAGPVVL----- 170
 Db 279 -----DSAGKSTEDTAYSPFKGG---YTKLEGESLGSAAAGSSSTLEPSTLSL 330
 QY 171 -----MAASFANAVYDRLPIDEEETPPRPT-----TGVOQQQPEASQSEVTSQAQ 218
 Db 331 YKSGALDEAAAYQSRDYNYNPLALAGPPPPPPHARIKLENP-LDYGSAWAAAAAQC 389
 QY 219 CESNLQGGNGGGGV-----AFYNLGMNMNPFQSGGDIYGMSSGGGGGG 264
 Db 390 RYGDLSLHAGAGAAGPGSGSPSAAASSWHTL-----FTAEGLYPCGGGGGGGG 442

```

RESULT 7
REF1_HUMAN
ID REF1_HUMAN STANDARD; PRT; 979 AA.
AC P22670;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MHC class II regulatory factor RFX1 (RFX) (Enhancer factor C) (EF-C).
GN RFX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=91071581; PubMed=2253877;
RA Reith W., Sanchez-Herrero C., Kober M., Silacci P., Berte C.,
RA Barras E., Mach B.;
RT "MHC class II regulatory factor RFX has a novel DNA-binding domain
RT and a functionally independent dimerization domain.";
RL Genes Dev. 4:1528-1540(1990).
RN [2]
IDENTITY BETWEEN RFX1 AND EF-C.
RP MEDLINE=94019311; PubMed=8413236;
RA Siegrist C.A., Durand B., Emery P., David E., Hearing P., Mach B.,
RA Reith W.;
RT "RFX1 is identical to enhancer factor C and functions as a
RT transactivator of the hepatitis B virus enhancer.";
RL Mol. Cell. Biol. 13:6375-6384(1993).
RN [3]
BINDING TO RPL30 PROMOTER.
RP MEDLINE=94040774; PubMed=8224874;
RA Safrany G., Perry R.P.;
RT "transcription factor RFX1 helps control the promoter of the mouse
RT ribosomal protein-encoding gene rpl30 by binding to its alpha
RT element.";
RL Gene 132:279-283(1993).
RN [4]
SHOWS THAT BLS IS NOT DUE TO RFX1.
RP MEDLINE=92375076; PubMed=1508204;
RA "The DNA-binding defect observed in major histocompatibility complex
RT class II regulatory mutants concerns only one member of a family of
RT complexes binding to the X boxes of class II promoters.";
RL Mol. Cell. Biol. 12:4076-4083(1992).
CC -1- FUNCTION: REGULATORY FACTOR ESSENTIAL FOR MHC CLASS II GENE
CC EXPRESSION. BINDS TO THE X BOXES OF MHC CLASS II GENES. ALSO BINDS
CC TO AN INVERTED REPEAT (ENH1) REQUIRED FOR HEPATITIS B VIRUS GENES
CC EXPRESSION AND TO THE MOST UPSTREAM ELEMENT (ALPHA) OF THE RPL30
CC PROMOTER.
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE RFX FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X58964; CAA41730.1;
CC EMBL: A20498; CAA01506.1;
CC PIR: A35913; A35913.
CC TRANSFAC: T0909;
CC TRANSFAC: T01673;
CC TRANSFAC: T01673;
CC MIM: 600006;
CC InterPro: IPR003150; RFX_DNA_binding.
CC Pfam: PF02257; RFX_DNA_binding; 1.
KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 381 411 GLY-RICH.
FT DNA_BIND 436 528 EXPERIMENTALLY DEDUCED.

```

```

FT DOMAIN 920 936 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 744 979 NECESSARY FOR DIMERIZATION.
SQ SEQUENCE 979 AA: 104728 MW; 556151F88C6AC9A2 CRC64;

Query Match 8.8%; Score 122; DB 1; Length 979;
Best Local Similarity 20.5%; Pred. No. 1;
Matches 78; Conservative 34; Mismatches 124; Indels 144; Gaps 14;

QY 17 PTGGATSSATASGSSRRPRGRPAGSKNK-----PKPPTIITRDSNVLRSVHL 66
|| || :|: || :|: || || || || || || || || || || || || || || ||
DB 44 PTAATPQPYVTELOSPQPGQXQYVTELPVAPAPSOPTGAPTPSPAPQYIVV 103
|| || || || || || || || || || || || || || || || || || || || ||
QY 67 EVTSGS-DISEAVS-----TYATRRGCGVCI---ISGTG-----AV 98
|| || || || || || || || || || || || || || || || || || || || ||
DB 104 TVSEGAMRASETSEASPGSTASQTGPTQVQVQVQGTQORLLVQTSVQAKPGHVSPLQL 163
|| || || || || || || || || || || || || || || || || || || || ||
QY 99 TNYTIQPAAPA-----GGGV-ITLHGRFDILS-----125
|| || || || || || || || || || || || || || || || || || || || ||
DB 164 TNIQVPOQALPTQRLVVQSAAPGKGGVSLTVHGTQQVHSPPPEQSPVOANSSSKTAGA 223
|| || || || || || || || || || || || || || || || || || || || ||
QY 126 LTGT-----ALPPPPAPP-AGGLTV-----144
|| || || || || || || || || || || || || || || || || || || || ||
DB 224 PTGTVPOQLQVHGVSQVPTQERSVVQATPQAPKPGVQPLTVQGLQPVHVAQEVQQLQ 283
|| || || || || || || || || || || || || || || || || || || || ||
QY 145 -----YLAGGQGVVGGVNVAGSLIASGPPVLMASFANAYVDRLPIDEEETP 191
|| || || || || || || || || || || || || || || || || || || || ||
DB 284 QVPVPHVYSSQVQVVEGGDASYTASAIRSSTYSYPETPLTYQTASTSYEAAATATQVST 343
|| || || || || || || || || || || || || || || || || || || || ||
QY 192 PPRITGVQO--QOPEASOSSEVTGS-----GAQACESNLGGNGGGGVAFYNLGMNNNFQ 245
|| || || || || || || || || || || || || || || || || || || || ||
DB 344 PATSQAVASSGSRPMYVSGVQVVASSTAGASNSGGSGGGGGGG-----391
|| || || || || || || || || || || || || || || || || || || || ||
QY 246 FSGDDIYMGSGGGGGGGGA 265
|| || || || || || || || || || || || || || || || || || || || ||
DB 392 -GGGGGGGGSGTGGGGGA 410
|| || || || || || || || || || || || || || || || || || || || ||

RESULT 8
ANDR_PAPHA STANDARD; PRT; 895 AA.
AC O97960;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Androgen receptor (dihydrotestosterone receptor).
GN AR OR NR3C4.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheciniae; Papio.
OX NCBI_TaxID=9557;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=98404153; PubMed=9732460;
RA Choong C.S., Kempainen J.A., Wilson E.M.;
RT "Evolution of the primate androgen receptor: a structural basis for
RT disease.";
RL J. Mol. Evol. 47:334-342(1998).
CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR3 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

CC -----
ENBL; U94176; AAC73047.1; -.
DR HSP; P06536; IGDC.
DR InterPro; IPR001103; Androgen_recep.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF021166; Androgen_recep; 1.
DR Pfam; PF001104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR000047; STROIDFINGER.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; ZnF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Steroid-binding.
FT DOMAIN 1 533 MODULATING (BY SIMILARITY).
FT FT 535 600 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 535 555 C4-TYPE.
FT ZN_FING 571 595 C4-TYPE.
FT DOMAIN 666 895 C4-TYPE.
FT DOMAIN 55 63 LIGAND-BINDING.
FT DOMAIN 69 74 POLY-GLN.
FT DOMAIN 178 182 POLY-GLN.
FT DOMAIN 357 366 POLY-PRO.
FT DOMAIN 381 387 POLY-ALA.
FT DOMAIN 434 448 POLY-GLY.
SQ SEQUENCE 895 AA; 96478 MW; 9020C0DC67F11E5D CRC64;

```

Query Match	8.7%	Score 121.5;	DB 1;	Length 895;
Best Local Similarity	23.4%	Pred. No. 0.99;		
Matches	70;	Conservative 34;	Mismatches 130;	Indels 65; Gaps 11;
Qy	1	MELNRSEADAKATTPTGGATSSATASGSSG--RRPRGPAGSKNKPPTTIITRDSP	58	
Db	174	MOLLOQQQOEAVS-----GSSGRAREASGAPTSCKDNYLGGTSTISDSA	219	
Qy	59	NVLRSHV-LEVTSSDISEAVSYTATRRG-CGVCIISGTGAVTNVTTQPAAPAGGGVIT	116	
Db	220	KELCKAVSVSMGLGVEALEHLSPEQURGDQMAPIVLGVPPVAPRPTCPALAECKGSL-	278	
Qy	117	LHGRFDILSTGTALPPPPGAGGLTVYLAGGQGVVGVNAGSLIASPVVL-----	170	
Db	279	-----DDSACKSTEDTAEYSPFKGG-----YTKLEGESLGCSSGAAGSSGTLELPSTLSL	330	
Qy	171	-----MAASFAANVYDRLPIDEEETPPORT-----TGVOOQOPEASQSSEVTGSAQA	218	
Db	331	YKSCALDEAAAYQSRDYNFPLALAGPPPPPPPHPHARIKLENP-LDYSANAAAAAQC	389	
Qy	219	CESNLQNGNGGGV-----AFYNLGMNMNMFQSGDDIYGMSSGGSGGGGG	264	
Db	390	RYGSLASLHGAGAGPGSGPSAAASSSWHTL-----FTAEESQLYCPGGGGGGGGG	442	

RESULT	9	ANDR_PANTR	ANDR_PANTR	STANDARD:	PRT:	911 AA.
AC	O97775;					
DT	30-MAY-2000	(Rel. 39, Created)				
DT	30-MAY-2000	(Rel. 39, Last sequence update)				
DT	30-MAY-2000	(Rel. 39, Last annotation update)				
DE	Androgen receptor (Dihydrotestosterone receptor).					
AR	OR NR3C4.					
OS	Pan troglodytes (Chimpanzee).					
OC	Eukaryota; Metazoa; Chordata;					
OC	Mammalia; Eutheria; Primates;					
OX	NCBI_TaxID=9598;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=98404153; PubMed=9732460;					
RA	Choong C.S., Kempainen J.A., Willson E.M.;					

```

RT Evolution of the primate androgen receptor: a structural basis for
RT disease.*;
RL J. Mol. Evol. 47:334-342(1998).
CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A A-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U94177; GAC73048.1; -.
DR HSSP; P06536; LGDC.
DR InterPro; IPR001103; Androgen_recep.
DR InterPro; IPR000536; Hormone_rec_llg.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF02166; Androgen_recep; 1.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROJDFFINGER.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; ZnF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Steroid-binding.
FT DOMAIN 1 549 MODULATING (BY SIMILARITY).
FT DNA_BIND 551 616 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 551 571 C4-TYPE.
FT ZN_FING 587 611 C4-TYPE.
FT DOMAIN 682 911 LIGAND-BINDING.
FT DOMAIN 57 78 POLY-GLN.
FT DOMAIN 84 88 POLY-GLN.
FT DOMAIN 192 196 POLY-GLN.
FT DOMAIN 371 380 POLY-PRO.
FT DOMAIN 395 401 POLY-ALA.
FT DOMAIN 448 464 POLY-GLY.
SQ SEQUENCE 911 AA; 998402 MW; 601B9BD4E697DAA4 CRC64;

```

Query Match	8.7%	Score	121.5	DB 1	Length	911
Best Local Similarity	23.4%	Pred. NO.	1			
Matches	70	Conservative	35	Mismatches	129	Indels
					65	Gaps
						11

QY	1	MELNRSEADAKAETPTTGATSSATASGSSG--RRPRGRPAGSKNKPPTTIITRDSP	58
		:	
		:	
		:	
Db	188	MLQLQQQQEAVSE-----GSSGRAREAGAPTSSKDNYLGGTSTISDSA	233
		:	
		:	
		:	
QY	59	NVLRSHV-LQVTSGDISEAVSYATRRG-CGVCIISGTGAVTNVITRQPAAPAGGGVIT	116
		:	
		:	
		:	
Db	234	KECKAVSVSMGLGVEALEHLSQCEURGOCMVAPLGVPPVAPRPTCAPLAECKGSL--	292
		:	
		:	
		:	
QY	117	LHGRFDLSTLTGTALPPAPPAGGLTVYLAGGGQVGVGNVAGSLTASGPVVL-----	170
		:	
		:	
		:	
Db	293	-----DDSACKSTEDTAEYSFFKGG---YTKGLEGESLGCSSNAAGSSGTLELPSTLSL	344
		:	
		:	
		:	
QY	171	-----MAASFANAVYDLRPIBEEETPPPT-----TGVOQQQPEASQSSEVTGSAQA	218
		:	
		:	
		:	
Db	345	YKSGALDEAAAYQSDYYNFPALAGPPPPPPPPHPPHARIKLENP-LDYCSAWAAAAAQC	403
		:	
		:	
		:	
QY	219	CESNLQGGNGGGV-----AFTNLGMNMNFPFSGGDIYMGSSGGSGGGGG	264
		:	
		:	
		:	
Db	404	RYGDPLASHCAGAAGPGSGSPSAASSSWITL-----FTAEICOLYPCPGGGGGGGGG	456
		:	
		:	
		:	

RESULT 10
ID ANDR HUMAN STANDARD; PRT; 919 AA.
AC P10275;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Androgen receptor (Dihydrotestosterone receptor).
GN AR OR NR3C4 OR DHTR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89112208; PubMed=3216866;
RA Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Migeon R.E.,
RA French F.S., Wilson E.M.;
RT "The human androgen receptor: complementary deoxyribonucleic acid
RT cloning, sequence analysis and gene expression in prostate.";
RL Mol. Endocrinol. 2:1265-1275(1988).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT CAIS MET-866.
RX MEDLINE=90083302; PubMed=2594783;
RA Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J.,
RA Wilson E.M., French F.S.;
RT "Sequence of the intron/exon junctions of the coding region of the
RT human androgen receptor gene and identification of a point mutation
RT in a family with complete androgen insensitivity.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258935; PubMed=2342476;
RA Govindan M.V.;
RT "Specific region in hormone binding domain is essential for hormone
RT binding and trans-activation by human androgen receptor.";
RL Mol. Endocrinol. 4:417-427(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=89017168; PubMed=3174628;
RA Chang C., Kokontis J., Liao S.;
RT "Structural analysis of complementary DNA and amino acid sequences of
RT human and rat androgen receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098909; PubMed=2911578;
RA Tilley W.D., Marcelli M., Wilson J.D., McPhaul M.J.;
RT "Characterization and expression of a cDNA encoding the human androgen
RT receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=91159943; PubMed=2293020;
RA Marcelli M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D.,
RA McPhaul M.J.;
RT "Definition of the human androgen receptor gene structure permits the
RT identification of mutations that cause androgen resistance: premature
RT termination of the receptor protein at amino acid residue 588 causes
RT complete androgen resistance.";
RL Mol. Endocrinol. 4:1105-1116(1990).
RN [7]
RP SEQUENCE OF 189-919 FROM N.A.
RX MEDLINE=88178111; PubMed=3353726;
RA Chang C., Kokontis J., Liao S.;
RT "Molecular cloning of human and rat complementary DNA encoding
RT androgen receptors.";
RL Science 240:324-326(1988).
RN [8]
RP SEQUENCE OF 468-919 FROM N.A.
RX MEDLINE=88240407; PubMed=3377788;
RA Trapman J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,
RA Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M.,
RA Mulder E., Brinkmann A.O.;
RT "Cloning, structure and expression of a cDNA encoding the human
RT androgen receptor.";
RL Biochem. Biophys. Res. Commun. 153:241-248(1988).
RN [9]
RP INTERACTION WITH RAN.
RX MEDLINE=99329028; PubMed=10400640;
RA Hsiao P.-W., Lin D.-L., Nakao R., Chang C.;
RT "The linkage of Kennedy's neuron disease to ARA24, the first
RT identified androgen receptor polyglutamine region-associated
RT coactivator.";
RL J. Biol. Chem. 274:20229-20234(1999).
RN [10]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=92220629; PubMed=1561105;
RA Sledzens H.F.B.M., Oostra B.A., Brinkmann A.O., Trapman J.;
RT "Trinucleotide repeat polymorphism in the androgen receptor gene
RT (AR).";
RL Nucleic Acids Res. 20:1427-1427(1992).
RN [11]
RP POLYMORPHISM OF POLY-GLY REGION.
RC TISSUE=Blood;
RA Lu J., Danielsen M.;
RT Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [12]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=97250535; PubMed=9096391;
RA Giovannucci E., Stampfer M.J., Krithivas K., Brown M., Brufsky A.,
RA Talcott J., Hennekens C.H., Kantoff P.W.;
RT "The CAG repeat within the androgen receptor gene and its
RT relationship to prostate cancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:3320-3323(1997).
RN [13]
RP ERRATUM.
RA Giovannucci E., Stampfer M.J., Krithivas K., Brown M., Dahl D.,
RA Brufsky A., Talcott J., Hennekens C.H., Kantoff P.W.;
RL Proc. Natl. Acad. Sci. U.S.A. 94:8272-8272(1997).
RN [14]
RP REVIEW ON VARIANTS
RX MEDLINE=93092459; PubMed=1458719;
RA Pinsky L., Trifiro M.A., Kaufman M., Beitel L.K., Mhatre A.,
RA Kazemi-Esfarjani P., Sabbaghian N., Lumbruso R., Alvarado C.,
RA Vasilidou M., Gottlieb B.;
RT "Androgen resistance due to mutation of the androgen receptor.";
RL Clin. Invest. Med. 15:456-472(1992).
RN [15]
RP REVIEW ON VARIANTS AIS.
RX MEDLINE=93339360; PubMed=8339746;
RA Brown T.R., Scherer P.A., Chang Y.-T., Migeon C.J., Ghirri P.,
RA Muroto K., Zhou Z.;
RT "Molecular genetics of human androgen insensitivity.";
RL Eur. J. Pediatr. 152 Suppl. 2:S62-S69(1993).
RN [16]
RP REVIEW ON VARIANTS.
RX MEDLINE=94059770; PubMed=8240973;
RA Sultan C., Lumbruso S., Poujol N., Belon C., Boudon C.,
RA Lobaccaro J.-M.;
RT "Mutations of androgen receptor gene in androgen insensitivity
RT syndromes.";
RL J. Steroid Biochem. Mol. Biol. 46:519-530(1993).
RN [17]
RP REVIEW ON VARIANTS.
RX MEDLINE=95023089; PubMed=7937057;
RA Patterson M.N., Hughes I.A., Gottlieb B., Pinsky L.;
RT "The androgen receptor gene mutations database.";
RL Nucleic Acids Res. 22:3560-3562(1994).
RN [18]
RP REVIEW ON VARIANTS.
RX MEDLINE=95352489; PubMed=7626493;
RA Brinkmann A.O., Jenster G., Ris-Stalpers C., van der Korput J.A.G.M.,

Bruggenwirth H.T., Boehmer A.L.M., Trapman J.;
 "Androgen receptor mutations."; *J. Steroid Biochem. Mol. Biol.* 53:443-448(1995).
 [19]
 REVIEW ON VARIANTS.
 RX MEDLINE-97169385; PubMed-9016528;
 RX Gottlieb B., Trifiro M.A., Lumbroso R., Vasiliou D.M., Pinsky L.;
 "The androgen receptor gene mutations database."; *Nucleic Acids Res.* 25:158-162(1997).
 [20]
 VARIANT LNCAP ALA-877.
 RX MEDLINE-91083633; PubMed-2260966;
 RX Veldscholte J., Ris-Stalpers C., Kuiper G.G.J.M., Jenster G.,
 RX Berrevoets C.A., Claassen E., van Rooij H.C.J., Trapman J.,
 RX Brinkmann A.O., Mulder E.;
 "A mutation in the ligand binding domain of the androgen receptor of
 human LNCAP cells affects steroid binding characteristics and
 response to anti-androgens."; *Biochem. Biophys. Res. Commun.* 173:534-540(1990).
 [21]
 VARIANTS CAIS CYS-774; GLN-831 AND MET-866.
 RX MEDLINE-91186983; PubMed-2082179;
 RX Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,
 RX Corfen J.L.;
 "Functional characterization of naturally occurring mutant androgen
 receptors from subjects with complete androgen insensitivity."; *Mol. Endocrinol.* 4:1759-1772(1990).
 [22]
 VARIANT CYS-774.
 RX MEDLINE-91310758; PubMed-1856263;
 RX Marcelli M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,
 RX McPhaul M.J.;
 "Androgen resistance associated with a mutation of the androgen
 receptor at amino acid 772 (Arg-->Cys) results from a combination of
 decreased messenger ribonucleic acid levels and impairment of
 receptor function."; *J. Clin. Endocrinol. Metab.* 73:318-325(1991).
 [23]
 VARIANT CAIS PRO-617.
 RX MEDLINE-91154385; PubMed-1999491;
 RX Marcelli M., Zoppi S., Grino P.B., Griffin J.E., Wilson J.D.,
 RX McPhaul M.J.;
 "A mutation in the DNA-binding domain of the androgen receptor gene
 causes complete testicular feminization in a patient with
 receptor-positive androgen resistance."; *J. Clin. Invest.* 87:1123-1126(1991).
 [24]
 VARIANT PAIS CYS-763.
 RX MEDLINE-91185626; PubMed-2010552;
 RX McPhaul M.J., Marcelli M., Tilley W.D., Griffin J.E.,
 RX Isidro-Gutierrez R.F., Wilson J.D.;
 "Molecular basis of androgen resistance in a family with a qualitative
 abnormality of the androgen receptor and responsive to high-dose
 androgen therapy."; *J. Clin. Invest.* 87:1413-1421(1991).
 [25]

Query Match 8.5%; Score 118.5; DB 1; Length 919;
Best Local Similarity 23.1%; Pred. No. 1.6;
Matches 69; Conservative 36; Mismatches 129; Indels 65; Gaps 11;

[illegible]

QY	171	-MAASFANVVDRLPPIEEETPPRT-----TGVOOQPPEASSSVTSGAQA	218
DG	346	YKSGALDEAAAYQSRYNFFPALAGAPPPPPPAPHRIKLNP-LDYGSAAAAAAC	404
QY	219	CESNLGGNGGGV-----AFYNLGNMNNFQFGSDIYHNSGGSGGGGG	264
DB	405	RVGDLASHLGAGPGSGPSAAAASSWHIL-----FTAEEQLYGPCGGGGGGGG	457
RESULT	11		
BET3_MESAU	ID	BET3_MESAU STANDARD; PRT;	367 AA.
OOC	O09029:		
DT	30-MAY-2000	(Rel. 39, Created)	
DT	30-MAY-2000	(Rel. 39, Last sequence update)	
DT	16-OCT-2001	(Rel. 40, Last annotation update)	
DE	BETA3 protein.		
OS	Mesocricetus auratus	(Golden hamster).	
OC	Eukaryota:	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OCC	Mammalia;	Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;	
OCC	Mesocricetus.		
OX	NCBI_TaxID=10036;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96140430; PubMed=8552091;		
RA	Peyton M., Stellrecht C.M.M., Maya F.J., Huang H.-P., Samora P.J., Tsai M.-J.;		
RT	"BETA3", a novel helix-loop-helix protein, can act as a negative regulator of BETA2 and MyoD-responsive genes.";		
RT	Mol. Cell. Biol. 16:626-633(1996).		
RRC	-!- FUNCTION: INHIBITS DNA BINDING OF TCF3 (E47) HOMODIMERS AND TCF3 (E47) / NEUROD1 HETERODIMERS AND ACTS AS A STRONG REPRESSOR OF NEURONAL AND MYOD-RESPONSIVE GENES, PROBABLY BY HETERODIMERIZATION WITH CLASS A BASIC HELIX-HLOOP-HELIX FACTORS. DESPITE THE PRESENCE OF AN INTACT BASIC DOMAIN, DOES NOT BIND TO DNA.		
OCC	-!- SUBUNIT: HETERODIMER WITH OTHER HLH PROTEINS, LIKE TCF3 (E47).		
OCC	-!- SUBCELLULAR LOCATION: Nuclear (Potential);		
OCC	-!- TISSUE SPECIFICITY: KIDNEY LUNG BRAIN AND PANCREAS (INSULINOMA).		
OCC	-!- SIMILARITY: BELONGS TO THE "BASIC HELIX-HLOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.		
OCC	-!		
OCC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isg-sib.ch/announce/) or send an email to license@sib-sib.ch .		
EMBL:	S80870; AAB50691.1; "		
InterPro:	IPIR003015; HLH_myc.		
IncrPro:	IPIR001092; HLH_dlm.		
Pfam:	PF00010; HLH; 1.		
SMART:	SM00353; HLH; 1.		
PROSITE:	PS00038; HELIX_LOOP_HELIX; 1.		
Nuclear protein:	Transcription regulation; Repressor.		
DOMAIN	11 14 POLY-ALA.		
FT	58 62 POLY-SER.		
FT	83 99 POLY-GLY.		
FT	174 179 POLY-GLY.		
FT	204 217 POLY-GLY.		
FT	229 240 BASIC DOMAIN		
FT	241 282 HELIX-HLOOP-HELIX MOTIF (BY SIMILARITY).		
FT	311 319 POLY-ALA.		
SEQ	SEQUENCE 367 AA; 35905 MW; GCABA9AFF96BE85F77 CRC64;		

Query Match 8.4%; Score 117; DB 1; Length 367;
Best Local Similarity 23.3%; Pred. No. 0.77;

QY 10 EAKAETPTGGATSSATASGSSGRRPRGRPRGAGSKNKPPTIITRDSPNVLRSHVLEVT 69

Db 34 EAAFRSTPPGMDLSTAPP-----PRRPASSSSSP----- 63
 QY 70 SGSDISEAVSYATRRGGVCIISTGAVTWTITROPAPAGAGGVITLHGRFDILSLTGT 129
 Db 64 -----LGC-----FEPADPEGAGLL----- 78
 QY 130 ALPPPPAPPAGGLTVYLAGGGGVVGNVAGSLIASGPVWLMAASFANAVYDRLP----- 184
 Db 79 -LPPG- -GGG- -AGGG- -GGGGGVSVPGLLVGSAGVGGDPNLSLPAGAAAL 127
 QY 185 -----IEE-----EETPPPTTG-----VOOOOPEASQSSEVTGSGNAQACE- 220
 Db 128 CLKYGESAGRGVSAESSGGEQSPDDSDGRCBLVLRAGGADPRASPGA- -GGGCTKVVVEG 185
 QY 221 ----SNLOGNG- -GGVAFVYNLGMNNFQFSGGDIYGMSSGGGGGGGAT 266
 Db 186 CSNAHLHGGAGLPPGGSTG-----SGG-----GGSGGGGGGGG 218
 RESULT 12
 CIKE_DROME STANDARD; PRT; 1174 AA.
 AC Q02280;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Potassium channel protein eag.
 GN EAG.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91262635; PubMed=1840699;
 RX Warmke J., Drysdale R., Ganetzky B.;
 RT "A distinct potassium channel polypeptide encoded by the Drosophila
 RT eag locus";
 RL Science 252:1560-1562(1991).
 CC -!- FUNCTION: PROTEIN EAG IS MOST PROBABLY A STRUCTURAL COMPONENT OF
 CC -!- THE POTASSIUM CHANNEL AND MEDIATES THE POTASSIUM PERMEABILITY OF
 CC -!- MEMBRANES.
 CC -!- SUBUNIT: DIMER (POTENTIAL).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- MISCELLANEOUS: THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND
 CC -!- IS CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT
 CC -!- EVERY THIRD POSITION.
 CC -!- MISCELLANEOUS: THE SEGMENT H5 IS THOUGHT TO LINE THE CHANNEL PORE.
 CC -!- SIMILARITY: TO THE MEMBERS OF THE POTASSIUM CHANNEL PROTEINS
 CC -!- OF THE SH SUPERFAMILY.
 CC -!- SIMILARITY: CONTAINS 1 CYCLIC NUCLEOTIDE-BINDING DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 PAS (PBR-ARNT-SIM) DIMERIZATION DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M61157; AAA28495.1; -;
 CC PIR; A40853; A40853.
 CC FLYBase; FBgn0000535; eag.
 CC InterPro; IPR000636; Cation_chan_non_lig.
 CC InterPro; IPR001622; Channel_pore_K.
 CC InterPro; IPR001610; PAC.
 CC InterPro; IPR000700; PAS-assoc_C.
 CC InterPro; IPR000014; PAS.
 CC InterPro; IPR000595; cNMP_binding.
 CC Pfam; PF00027; cNMP_binding; 1.

DR Pfam; PF00520; ion_trans; 1.
 DR DR PF00785; PAC; 1.
 DR SMART; SM00086; PAC; 1.
 DR SMART; SM00091; PAS; 1.
 DR SMART; SM00100; cNMP; 1.
 DR PROSITE; PS00042; cNMP_BINDING_3; 1.
 DR PROSITE; PS01112; PAC; 1.
 DR PROSITE; PS01113; PAS; 1.
 KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KW Glycoprotein; Phosphorylation; Repeat.
 FT TRANSMEM 227 246 SEGMENT S1 (POTENTIAL).
 FT TRANSMEM 269 291 SEGMENT S2 (POTENTIAL).
 FT TRANSMEM 314 335 SEGMENT S3 (POTENTIAL).
 FT TRANSMEM 343 369 SEGMENT S4 (POTENTIAL).
 FT TRANSMEM 372 393 SEGMENT S5 (POTENTIAL).
 FT TRANSMEM 442 467 SEGMENT H5 (POTENTIAL).
 FT TRANSMEM 471 493 SEGMENT S6 (POTENTIAL).
 FT DOMAIN 43 97 PAS.
 FT NP_BIND 113 165 CNMP.
 FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1174 AA; 126236 MW; 344C8DC06E4340E CRC64;
 Query Match 8.4%; Score 117; DB 1; Length 1174;
 Best Local Similarity 21.6%; Pred. No. 2.5;
 Matches 58; Conservative 23; Mismatches 94; Indels 94; Gaps 8;
 QY 20 GATSSATASSSSRRPRGRP-----AGSKNKPKPTITRDSPNVLRSHVLEVT 70
 Db 932 GASSGGNAPDNNSSGQTTPGDEICAGCGAGGGTPTTQAPPT----- 972
 QY 71 GSDISEAVSYATRRGGVCIISTGAVTWTITROPAPAGAGGVITLHGRFDILSLTGT 130
 Db 973 -SANTSPVDVTITISSPGA---SGSGSCTGAGAGSAGAGGAGGLDPGATVVSAGGNG 1028
 QY 131 LPP-----PAPGAGGLTVYLAGGGGVVGNVAGS-LIASGPVWLMAASFA 176
 Db 1029 LGPLMLKRRRSKSGKAPAPPEQ---TLASTAGTATAAPAGVAGSGMTSSAP----- 1076
 QY 177 NAVYDRLPIDEEETPPPTTGVCQQQPEASQSSEVTGSGAACESNLQGGNGGGVAFYN 236
 Db 1077 -----ASADQQQHQSRADQSPPTFGAEL----- 1100
 QY 237 LGMNNNFQFSGGDIYGMSSGGGGGGA 265
 Db 1101 LHLRLLEEDFTAAQLPSTSSGGAGGGGGS 1129
 RESULT 13
 IRS1_MOUSE STANDARD; PRT; 1233 AA.
 AC P35569;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Insulin receptor substrate-1.
 GN IRS1 OR IRS-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94220494; PubMed=8167159;
 RA Araki E., Haag B.L. III, Kahn C.R.;
 RT "Cloning of the mouse insulin receptor substrate-1 (IRS-1) gene and
 RT complete sequence of mouse IRS-1";
 RL Biochim. Biophys. Acta 1221:353-356(1994).
 RN [2]
 RP SEQUENCE FROM N.A.

7	EADKAETPTTGATSSATASGSSGRPRGPAGSKNKPPIITRDSPNVLSHVL	66
:	:	:
59	ESDHSNDHHQQRDPDNPNTSSAPCKRPGRPPGSKNKAKPPIIVTRDSPNALRSHVL	118
:	:	:
67	EVTSGSDISEAVSYATRRCGGCIISFGAVTNTIROPAA-----AGGGVITLHG	119
:	:	:
119	EVSPGADVSESVYARRRGVRGVSLGGNGTVSNVTLRQPYPNGCGVGSCGGGVWTLHG	178
:	:	:
120	RFDILSLGTALPPPAPPAGGLTYVLAGQGQGVVGVGNVAGSLIASGVPVILMAASFANAV	179

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:19:08 ; Search time 103.68 Seconds
(without alignments)
450.508 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNSEADEAKETPTGG.....IYMSGSGGGGGATPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_prodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1392	100.0	270	10	O65489	O65489 arabidopsis
2	706.5	50.8	311	10	O9S7C9	Q9S7C9 arabidopsis
3	646.5	46.4	292	10	O23620	O23620 arabidopsis
4	645.5	46.4	302	10	O9C9K7	Q9C9K7 arabidopsis
5	637.5	45.8	315	10	O9SR17	Q9SR17 arabidopsis
6	629.5	45.2	317	10	O22130	O22130 arabidopsis
7	625.5	44.9	339	10	O9S270	Q9S270 arabidopsis
8	616	44.3	285	10	O82166	O82166 arabidopsis
9	605	43.5	324	10	O49662	O49662 arabidopsis
10	581	41.7	265	10	O9LZX7	Q9LZX7 arabidopsis
11	572	41.1	310	10	O9M2S3	Q9M2S3 arabidopsis
12	554	39.8	257	10	O9S7G4	Q9S7G4 arabidopsis
13	499.5	35.9	276	10	O9LTA2	Q9LTA2 arabidopsis
14	417	30.0	206	10	O9M9R4	O9M9R4 arabidopsis
15	407	29.2	265	10	O94LY9	Q94LY9 oryza sativ
16,	289.5	20.8	404	10	Q9LVB0	Q9LVB0 arabidopsis

17	289	20.8	771	10	Q9XHV4	O9xhv4 oryza sativ
18	283	20.3	405	10	Q9AYM0	Q9aym0 oryza sativ
19	282	20.3	351	10	Q94F52	Q94f52 arabidopsis
20	274	19.7	296	10	O22812	O22812 arabidopsis
21	274	19.7	351	10	O23142	O23142 arabidopsis
22	264.5	19.0	419	10	Q9FHM5	Q9fhm5 arabidopsis
23	259	18.6	826	10	Q9SH33	Q9sh33 arabidopsis
24	258.5	18.6	300	10	Q9ZRR7	Q9zrr7 antirrhinum
25	255.5	18.4	334	10	O49658	O49658 arabidopsis
26	248.5	17.9	386	10	O49FIR1	O49f96 pisum sativ
27	246.5	17.7	347	10	O04696	O04696 pisum sativ
28	243	17.5	439	10	O49350	O49350 arabidopsis
29	243	17.5	439	10	O49350	O49350 arabidopsis
30	242.5	17.4	334	10	Q94010	Q94010 arabidopsis
31	239.5	17.2	404	10	O04695	O04695 pisum sativ
32	239	17.2	455	10	O04694	O04694 arabidopsis
33	227	16.3	365	10	Q9S273	Q9sz73 arabidopsis
34	225	16.2	348	10	O80834	O80834 arabidopsis
35	221	15.9	418	10	Q9SR16	Q9sr16 arabidopsis
36	218	15.7	345	10	O81321	O81321 arabidopsis
37	211.5	15.2	348	10	O9M2D3	O9m2d3 arabidopsis
38	162.5	11.7	574	10	Q9SJO2	Q9sjq2 arabidopsis
39	148	10.6	848	5	Q9V889	Q9v889 drosophila
40	146	10.5	405	5	Q23057	Q23057 caenorhabdi
41	143.5	10.3	818	6	Q9N1P0	Q9nlp0 bos taurus
42	139	10.0	1468	5	Q9CUB5	Q9cub5 galliera me
43	138.5	9.9	265	5	Q23347	Q23347 caenorhabdi
44	137	9.8	694	16	O53212	O53212 mycobacteri
45	136.5	9.8	1610	16	Q92KQ8	Q92kq8 rhizobium m

ALIGNMENTS

RESULT 1

O65489 O65489 PRELIMINARY; PRT; 270 AA.
AC O65489;
DT 01-AUG-1998 (TREMREL. 07, Created)
DT 01-AUG-1998 (TREMREL. 07, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE HYPOTHETICAL 26.8 KDA PROTEIN.
GN F23E12.50 OR AT4G35390.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_taxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Hoheisel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.F.X.,
RA Schueller C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022604; CAA18730.1; -;
DR EMBL; AL161587; CAB80256.1; -;
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 26774 MW; 4C21487C2719E114 CRC64;

Query Match 100.0%; Score 1392; DB 10; Length 270;
Best Local Similarity 100.0%; Pred. No. 4.7e-86;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 236 -----QQQQ-----LGGGNG-----GGNLFPEVAAGGGGLPFFNLPNMOPNV 275
QY 245 OFSGDIYMGSGGGGG 262
Db 276 QL---PVEGWPGNSGGRG 290

RESULT 4
Q9C9K7 PRELIMINARY; PRT; 302 AA.
AC Q9C9K7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 30.6 KDA PROTEIN.
GN F14G6.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., S.Y.,
RA White O., Alonso J., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Buehler E., Chan A., Chao Q., Conway A.R., Creasy T.H., Dewar K.,
RA Chung M.K., Conn L., Luan J., Nelson C.R., Nierman W.C., Osborne B.I.,
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kvan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408:816-820(2000).
RE EMBL; AC015450; AAGS1949.1; -.
DR InterPro: IPR000637; AT_hook.
DR Pfam: PF02178; AT_hook; 1.
DR SMART: SM00384; AT_hook; 1.
KW Hypothetical protein.
SQ SEQUENCE 302 AA; 30615 MW; 6EC755021E3DD375 CRC64;

Query Match 46.4%; Score 645.5; DB 10; Length 302;
Best Local Similarity 49.5%; Pred. No. 7.2e-37;
Matches 138; Conservative 33; Mismatches 63; Indels 45; Gaps 7;

QY 9 DEAKAETPTTGGATSSATAGSSGRRPRGPKGKPKPTTIITRDSNVLRSHVLEV 68
Db 52 DESDNKDP-----GSDPVTSGSTGKPRGPRGPKGKPKPPVITRDSNVLRSHVLEV 106
QY 69 TSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIROPAPAG-----GGVITLHGRF 121
Db 107 SSGADIVESVTVYARRRGRGVSIISGNGTVANVSLRQPATTAHANGNGTGGVVALHGRF 166
QY 122 DILSLTGTALPPPPAGGAGLTIVLAGGQGVGGNVAGSLIAGSPVVLMAASFANAVYD 181
Db 167 EILSLTGTIVLPPPPAGGSLIFSLGQGVIGGNVAVPLVAGSPVILMAASFNSATFE 226
QY 182 RLPTEEETPPRTTGVQQQPEASQSSEVTGSGAQACESNLGGNGG-----GG 231
Db 227 RLPLEDE-----GGEGEGEGEV-GBGG-----GGEGGPPPATSSPPSG 264
QY 232 VAFYNLGNMNMNFQSGGDIYMGSGGGGGGATRPAP 270

```

```

Db 265 AGOQLRGNMSGYDQFAGDPH-LLGWGAAAAAAPPAPAF 302
RESULT 5
Q9SRI7 PRELIMINARY; PRT; 315 AA.
AC Q9SRI7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F7018.4 PROTEIN.
GN F7018.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F7018 genomic sequence."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC011437; AAF04888.1; -.
SQ SEQUENCE 315 AA; 32037 MW; 1766141EC657C448 CRC64;

Query Match 45.8%; Score 637.5; DB 10; Length 315;
Best Local Similarity 50.0%; Pred. No. 2.7e-36;
Matches 132; Conservative 37; Mismatches 58; Indels 37; Gaps 5;

QY 17 PTGGATSSATAGSSGRRPRGPKGKPKPTTIITRDSNVLRSHVLEVTSQSDISE 76
Db 70 PREGAVEAPT-----RRPRGPKGKPKPTTITRDSNVLRSHVLEVTSQSDISE 122
QY 77 AVSTYATRRGCGVCIISGTGAVTNVTIROP-----AAPAGGVITLHGRFDILSLTGT 130
Db 123 TLATFARRRGICILSGTGVNTVLRQPTAAVAAAPGAALALQGRFELSLTGSF 182
QY 131 LPPAPPGAGGLTVLAGGQGVGGNVAGSLIAGSPVVLMAASFANAVYDRLPTEEBET 190
Db 183 LPPAPPGAGGLTVLAGGQGVGGNVAGSLIAGSPVVLMAASFANAVYDRLPTEEBET 242
QY 191 PP-----PRTTGVQQQPEASQSSEVTGSGAQACESNLGGNGGQGVAFYNLGNMNMNFQ 246
Db 243 AERGGGGSGGVVPGQ-----LGGGSGPLSSGAGGDNQGLPVYNNMFGNLVS--- 290
QY 247 SGGDIYMGSGGGGGGGGATRPAP 270
Db 291 -----NGSGGGGQMSQGEAY 306

RESULT 6
Q22130 PRELIMINARY; PRT; 317 AA.
AC Q22130;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE AT-HOOK DNA-BINDING PROTEIN.
GN AT2G45430.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;

```

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana";
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC002387; AAB82621.1; -.
 KW DNA-binding.
 SQ SEQUENCE 317 AA; 33519 MW; 3A3677B991AE25F0 CRC64;

Query Match 45.2%; Score 629.5; DB 10; Length 317;
 Best Local Similarity 54.1%; Pred. No. 9.5e-36;
 Matches 126; Conservative 37; Mismatches 47; Indels 23; Gaps 5;

QY 2 ELNRSEADKA-ETPTTGGATSSATSSGSSGRRPRGRGPKGKPKPTTIITRDSN 59
 Db 55 DIDPNESSAGKQSTPGSGESGGGDNHITRRPRGRGPKGKPKPTTIITRDSN 114
 QY 60 VLRSVLEVTSGSDISEAVSYATRRGCGVCIISGTGAVNTVITROPAPAGGG--VITL 117
 Db 115 ALKSHVMEVANGCDVMESVTFARRRGICVLISGNGAVNTVITROPASVPGGGSVVNL 174
 QY 118 HGRFDLSLTGTALPPPPAGGGLTVYLAGGGQGVGVGNVAGSLIASGPPVLMASFPAN 177
 Db 175 HGRFELSLSGFLPPPPAPASGLITVYLAGGGQGVGVGNVAGSLIASGPPVLMASFPAN 234
 QY 178 AVYDRLPFIEETPTTGGVQQOQPEASQSEVTSVTSQAQACSNLQGGNGGG 230
 Db 235 AAYERLPLEEDD-----QEEQTAGAVANNIDGNATM-----GGG 268

RESULT 7
 Q9S270 PRELIMINARY; PRT; 339 AA.
 AC Q9S270;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE PUTATIVE DNA-BINDING PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 GN F16J13.120 OR AT4G12050.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL049638; CAB40946.1; -.
 DR EMBL; AL161533; CAB78248.1; -.
 KW DNA-binding.
 SQ SEQUENCE 339 AA; 35785 MW; A7E3A9C73B6BBE21 CRC64;

Query Match 44.9%; Score 625.5; DB 10; Length 339;
 Best Local Similarity 49.2%; Pred. No. 1.9e-35;
 Matches 129; Conservative 28; Mismatches 48; Indels 57; Gaps 3;

QY 4 NRSEADKAETPTTGGATSSATSSGSSGRRPRGRGPKGKPKPTTIITRDSNVLRS 63
 Db 88 NTNSGSEKEMSLHGGGSGEGQETRRPRGRGPKGKPKPTTIITRDSNALRT 147
 QY 64 HVLEVTSGSDISEAVSYATRRGCGVCIISGTGAVNTVITROPAPAGGGVITLHGRFDI 123
 Db 148 HVMEIGDGCIDVCMATFARRRGVCMVSGTGVNTVITROPFGSP-GSVSLHGRFEI 206
 QY 124 LSLTGTALPPPPAGGGLTVYLAGGGQGVGVGNVAGSLIASGPPVLMASFPANVYDRL 183
 Db 207 LSLSGSFLPPPPAPPAATGLSVYLAGGGQGVGVGNVAGSLIASGPPVLMASFPANVYDRL 266
 QY 184 PISEETPTTGGVQQOQPEASQSEVTSVTSQAQACSNLQGGNGGGVAFYNLGNMNN 243
 Db 267 PLEEDDQTP-----VQGGGGGG----- 285
 QY 244 FQFSGGDIYMGSGSGSGGGGGA 265
 Db 286 -----GGGGMGS 292

RESULT 8
 O82166 PRELIMINARY; PRT; 285 AA.
 AC O82166;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PUTATIVE AT-HOOK DNA-BINDING PROTEIN.
 GN AT2G35270.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana";
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004667; AAC61811.1; -.
 KW DNA-binding.
 SQ SEQUENCE 285 AA; 29144 MW; B7BC11E3732CB968 CRC64;

Query Match 44.3%; Score 616; DB 10; Length 285;
 Best Local Similarity 49.6%; Pred. No. 7.1e-35;
 Matches 130; Conservative 30; Mismatches 64; Indels 38; Gaps 6;


```
Qy 11 AKAEPTTGGATSSATASGSSG---RRPRGRPAGSKNPKPPTIITRSPNVLRSHVL 66
Db 57 ASNDGSLGG-----GGGGGDLVVRPRGRPAGSKNPKPVPVIVRESANTLRAHIL 110
Qy 67 EVTSGSDISEAVTYATRRGCGVCIISCTGAVTNTVIRPAAAGGCVITLHGRFDILSL 126
Db 111 EVGSGDVFECISTYARRRQICVLSTGTVTNVSIQPT--AAGAVVTLRGTFEILSL 168
Qy 127 TGTALPPPPAGCAGLVYLAGGCGVVGNGVAGSLIASGPPVLMASAFNAVYDRLPIE 186
Db 169 SGSELPAPPAGGSLTIFLAGAQQVVGNGVAGELMAAGPVMVMAASFNTNAVYERLPLD 228
Qy 187 EETPPPTTGVQQQPEASQSEVTSVTSVTSVTSVTSVTSVTSVTSVTSVTSVTSVTSV 246
Db 229 EHE-----EHLQSGGGGGGNNYSEAT-----GGGGGLPFFNLPMSPQI-- 268
Qy 247 SGGDIYGMGGGGGGGATRP 268
Db 269 -----GVESQGNHAGAGRAP 284

RESULT 9
O49662 ID O49662 PRELIMINARY; PRT; 324 AA.
AC O49662;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE DNA BINDING PROTEIN.
GN T12H17.200 OR AT4G22810.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K., Schueller C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021635; CAAL16566.1; -.
DR EMBL; AL161558; CAB79236.1; -.
SQ SEQUENCE 324 AA; 34339 MW; B8EF7520B474C528 CRC64;

Query Match 43.5%; Score 605; DB 10; Length 324;
Best Local Similarity 56.8%; Pred. No. 4.6e-34;
Matches 129; Conservative 26; Mismatches 42; Indels 30; Gaps 6;

Qy 19 GCAATSSATASGSSG-----RRPRGRPAGSKNPKPPTIITRSPNVLRSHVLEVTSGSD 73
Db 88 GG---SGEGGGGGDQHMTRPRGRPAGSKNPKPPTIITRSDANALRTHVMEIGDGD 144
Qy 74 ISEAVTYATRRGCGVCIISCTGAVTNTVIRPAA--PAGGCVITLHGRFDILSLTGTALP 132
Db 145 LVESVATFARRRGVCVMSGTGNTVNTVIRPQSGHSPGVSLSHGRFEILSLSGSFLP 204
Qy 133 PPAPGAGLVYLAGGCGVVGNGVAGSLIASGPPVLMASAFNAVYDRLPIEEEE--T 190
Db 205 PPAPPTATLSVYLAGGCGVVGNGVAGSLIASGPPVLMASAFNAVYDRLPIEEEE--T 264
Qy 191 P-----PPTTGVQQQPEASQSEVTSVTSVTSVTSVTSVTSVTSVTSVTSVTSVTSV 225
Db 265 PVHGGGGGSLSPMMQQLHQQQAM-----SCHQGLPPNLLG 304
```

```
RESULT 10
Q9LZX7 ID Q9LZX7 PRELIMINARY; PRT; 265 AA.
AC Q9LZX7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 28.4 KDA PROTEIN.
GN T4C21_280.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cholsene N., Robert C., Brottier P., Wincker P., Cattolico L.,
RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Rudd S.,
RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162295; CAB82691.1; -.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28412 MW; 31162D06E4BAD0AD CRC64;

Query Match 41.7%; Score 581; DB 10; Length 265;
Best Local Similarity 50.2%; Pred. No. 1.6e-32;
Matches 119; Conservative 31; Mismatches 57; Indels 30; Gaps 3;

Qy 34 RRPRGRPAGSKNPKPPTIITRSPNVLRSHVLEVTSGSDISEAVTATRRGCGVCIIS 93
Db 59 RRPRGRPAGSKNPKPPTIITRSDANAFCHVMEITNACDVMSLAVFARRRQGVCLT 118
Qy 94 GTGAVTNTVIRPAAAGGCVITLHGRFDILSLTGTALPPPPAGGAGLVYLAGGCGOV 153
Db 119 GNGAVNTVIRPQ---GGGVVSLHGRFEILSLSGSFLPPPPAASGLKVLVLAGGQGV 174
Qy 154 VGNVAGSLIASGPPVLMASAFNAVYDRLPIEEEEETPPPTTGVQQQPEASQSEVTS 213
Db 175 IGGSVVGLPTASSPVVMAAFNAGSYERLPLEEEE-----ETREIDG 218
Qy 214 SGAAACESNLQGGGGGVAFYNLGMNMFQSGDIIYGMGGGGGGGATRPAPF 270
Db 219 NAARLGTQTQQLMQDATSFIGSPSNLINSVSLGEAY-----WGTRQPSF 265

RESULT 11
Q9M2S3 ID Q9M2S3 PRELIMINARY; PRT; 310 AA.
AC Q9M2S3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 32.6 KDA PROTEIN.
GN T22E16.220.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Benes V., Wurmback E., Drzonek H., Ansorge W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
```